4. Análisis multivariable: métodos descriptivos más utilizados comúnmente

Los métodos de análisis multivariable se utilizan para estudiar las relaciones que hay entre más de dos variables.

4.1. Análisis factorial

4.1.1. Cuándo tenemos que utilizar el análisis factorial

El análisis factorial forma parte del conjunto de métodos de análisis multivariable cuyo objetivo consiste en estudiar las relaciones de interdependencia que se producen entre un conjunto de variables o individuos.

Se utiliza cuando queremos resumir la información que contiene una matriz de datos individuos/variables, tal como se muestra en el ejemplo, reemplazando las variables iniciales por un número menor de variables compuestas o factores, y perdiendo el mínimo posible de la totalidad de la información que contienen las variables iniciales.

Las matrices de datos de individuos por variables se explican en el subapartado 2.1 de este módulo didáctico.

Ejemplo de matriz de datos individuos por variables

En una encuesta realizada a una muestra de 1.000 estudiantes, se les pidió que valoraran en una escala de 1 a 6, en la cual 1 significaba “nada deseable” y 6 “totalmente deseable”, lo deseable que les parecía una serie de veinte características referentes al comportamiento de sus profesores. Las características son las siguientes:

V1 Tratan a todos los estudiantes aproximadamente igual.
V2 Incitan a los estudiantes a preguntar.
V3 Próximos, cercanos a los estudiantes.
V4 Muestran interés por los problemas de los estudiantes.
V5 Poseen un carácter agradable.
V6 Dan explicaciones que provocan un gran interés.
V7 Utilizan métodos propios para facilitar la instrucción.
V8 Visten con gusto, van limpios y aseados.
V9 Son puntuales.
V10 Son tranquilos, tienen control de sí mismos.
V11 Parecen seguros de sí mismos.
V12 Son ordenados en sus explicaciones.
V13 Encaran el lado bueno de las cosas.
V14 Desarrollan la asignatura con un sentido crítico.
V15 Tienen respuestas ingeniosas y completas.
V16 Dan explicaciones adaptadas a la realidad.
V17 Son tolerantes con el error de los estudiantes.
V18 Saben apreciar los esfuerzos del estudiante.
V19 Tienen buena dicción, son plenamente audibles.
V20 Saben controlar una situación de nerviosismo.

La matriz de resultados obtenidos fue la siguiente:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
Individuo 1	2	3	2	3	2	1	4	1	2	2	3	2	3	2	1	4	1	2	2	3
Individuo 2	2	4	5	2	2	5	4	2	2	2	4	5	2	2	5	4	2	2	5	2
Individuo 3	4	5	4	2	4	2	3	5	4	4	5	4	2	4	2	3	5	4	4	2
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Individuo 1000	1	4	2	3	5	6	1	3	1	3	4	2	3	1	6	5	3	1	2	3

En el ejemplo anterior, ¿es necesario guardar los 20.000 valores que hemos obtenido o bien podemos sintetizar (resumir) toda esta información en una, dos o tres variables compuestas? ¿No hay una relación determinada entre las variables iniciales y, por lo tanto, no podemos eliminar algunas que tan sólo aportan una información marginal y poco interesante? Si hay una relación de interdependencia sistemática en el conjunto de las variables, ¿no puede ser debida originalmente a algunos factores más fundamentales (latentes)? ¿No podemos considerar las variables originales como simples índices de estos factores fundamentales?

Es a este tipo de preguntas al que pretende responder el análisis factorial; como es lógico, no hay una respuesta única a preguntas de este tipo y, por lo tanto, se ha propuesto una gran cantidad de definiciones. Eso nos lleva a considerar el análisis factorial no como un método único, sino como una familia de métodos.

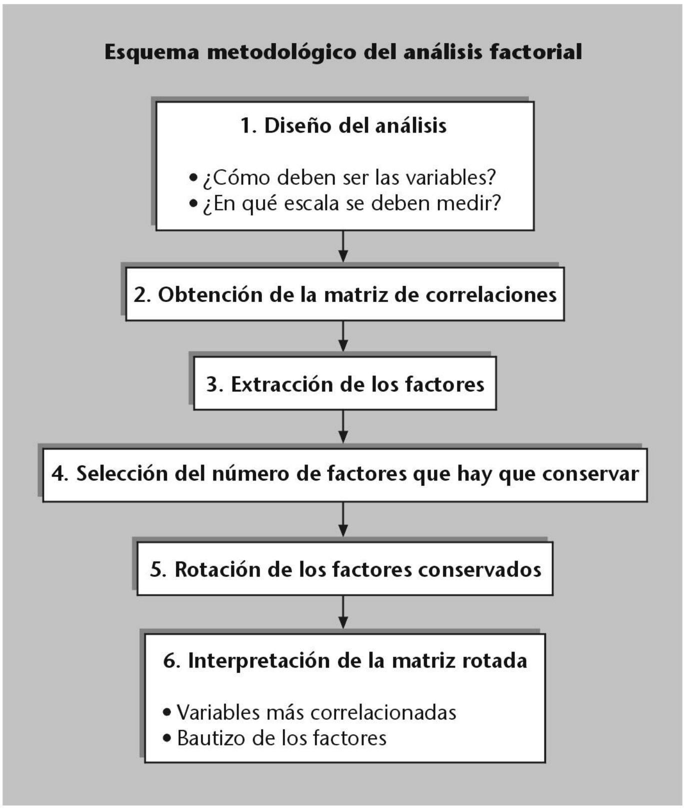
> El objetivo principal del análisis factorial consiste en identificar la estructura de las relaciones entre variables o entre individuos.

Si el objetivo de la investigación consiste en sintetizar las variables, el análisis recibe el nombre de **análisis factorial tipo *R***. En cambio, si el objetivo consiste en buscar los factores en el espacio de los individuos, lo cual permite agrupar a los individuos que tienen comportamientos análogos en relación con las variables sobre las cuales se lleva a cabo el análisis, el análisis recibe el nombre de **análisis factorial tipo Q.** !

A continuación, nos centraremos en el análisis factorial tipo R, por el hecho de que se trata del más utilizado en investigación comercial.

4.1.2. Metodología del análisis factorial

El proceso metodológico del análisis factorial consta de seis etapas:



Etapa 1: diseño del análisis factorial

El diseño del análisis hace referencia al tipo de variables que hay que utilizar y a las escalas de medida en las cuales deben ser medidas.

1) Las variables que se utilizan en el análisis deben ser comparables; corremos el riesgo de que los factores obtenidos no tengan ningún sentido en caso de que se introduzcan al mismo tiempo en el análisis variables socioeconómicas, variables de actitudes, variables de comportamiento, etc.

2) Si las escalas sobre las cuales se han medido las variables son muy diferentes, hay que **normalizarlas**, ya que, como veremos, la varianza de cada variable interviene en el análisis y las variables que presenten una mayor varianza tendrían unas ventajas determinadas. Normalizar las variables equivale a centrarlas y a reducirlas. Si X_p es la variable inicial, la variable X'_p normalizada será:

$$X'_p = \frac{X_p - \overline{X}}{S_p},$$

donde $\overline{X}$ es la media de la variable, y S_p, su desviación típica.

3) Por último, las variables introducidas deben medirse en escalas cuantitativas, es decir, métricas o de intervalo, las cuales no son precisamente las más abundantes en investigación de mercados.

Etapa 2: obtención de la matriz de correlaciones

A partir de la matriz de datos inicial, obtenemos la matriz de correlaciones entre variables.

Ejemplo

En nuestro ejemplo, a partir de la matriz de datos inicial expuesta más arriba, donde cada una de las filas de la matriz indica la valoración de cada estudiante sobre cada una de las veinte características propuestas, obtenemos la matriz de correlaciones entre variables siguiente:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
V1	1.00																		
V2	.37	1.00																	
V3	.77	.73	1.00																
V4	.86	.64	.96	1.00															
V5	.83	.40	.86	.86	1.00														
V6	.14	.74	.49	.46	.24	1.00													
V7	.15	.81	.52	.49	.19	.96	1.00												
V8	.42	.49	.53	.42	.51	.42	.42	1.00											
V9	.59	.38	.50	.48	.30	-.02	.05	.23	1.00										
V10	.36	.46	.63	.57	.70	.59	.49	.59	-.15	1.00									
V11	-.09	.61	.25	.17	.07	.84	.77	.29	-.20	.63	1.00								
V12	.14	.19	.16	.10	.24	.34	.18	.14	.09	.51	.63	1.00							
V13	.58	.66	.86	.79	.84	.61	.54	.57	.20	.89	.56	.51	1.00						
V14	-.13	.63	.22	.15	.06	.83	.76	.21	-.24	.56	.97	.60	.52	1.00					
V15	-.04	.66	.30	.24	.10	.91	.85	.30	-.15	.60	.97	.59	.56	.97	1.00				
V16	.12	.61	.40	.39	.17	.81	.80	.16	-.01	.55	.82	.54	.53	.83	.88	1.00			
V17	.96	.44	.87	.91	.90	.15	.16	.40	.57	.43	-.08	.13	.68	-.08	-.02	.15	1.00		
V18	.90	.54	.90	.90	.88	.13	.17	.42	.64	.43	-.05	.10	.70	-.06	-.02	.12	.96	1.00	
V19	.08	.62	.37	.30	.27	.80	.70	.34	-.13	.73	.96	.75	.69	.94	.94	.83	.11	.11	1.00
V20	-.21	.67	.26	.15	.07	.78	.73	.22	-.24	.59	.94	.51	.54	.96	.93	.77	-.11	-.04	.89

Matriz de correlaciones entre variables.

Etapa 3: extracción de los factores

El paso siguiente consiste en obtener, a partir de la matriz de correlaciones, los factores que identifiquen la estructura subyacente de las relaciones entre las variables iniciales. Con esta finalidad, el método más utilizado en investigación comercial es el **análisis factorial de componentes principales**.

Esta técnica puede resumirse como un método en el que se transforman las variables originales en unas nuevas variables que son una combinación lineal de las variables iniciales y que, además, no están correlacionadas. Estas nuevas variables se llaman **factores** o **componentes principales**.

El método busca restituir la máxima cantidad de información posible contenida en las variables iniciales en un número mínimo de factores. La medida de la cantidad de información restituida por cada componente principal es la **varianza**. Por este motivo, los factores se obtienen según la varianza restituida por cada uno. Así, el componente principal que se obtiene en primer lugar es el que restituye la mayor proporción de la varianza contenida en la matriz de correlaciones; el segundo factor es el que, de la varianza restante, restituye el mayor porcentaje, y así sucesivamente hasta llegar al último factor, que restituye la menor cantidad de varianza inicial.

Si tenemos:

- P variables iniciales $p = 1, ..., P$,
- I individuos $i = 1, ..., I$,

podemos extraer K factores; siendo $K = \min(P, I) - 1$.

Los factores extraídos tienen que cumplir las tres condiciones siguientes: !

1) **Linealidad**: cada factor es una combinación lineal de las variables iniciales.

$$F_k = a_{K1}X_1 + a_{k2}X_2 + ... + a_{kp}X_p$$

donde:

F_k = el k-ésimo factor;

a_{kp} = el coeficiente de la variable X_p en el factor F_k;

X_p = los valores de la variable p tomados por los I individuos.

2) **Independencia**: los factores no están correlacionados entre sí.

$$\mathrm{Corr}(F_k; \mathrm{F}_{k'}) = 0 \qquad \forall k \neq k'$$

3) **Varianza restituida por cada factor:** el primer factor restituye la proporción más alta de varianza contenida en la matriz de correlaciones; el segundo factor, la segunda; el tercero, la tercera; el cuarto, la cuarta, y así sucesivamente.

$$\mathrm{Var}\ F_1 \geq \mathrm{Var}\ F_2 \geq ... \geq \mathrm{Var}\ F_k$$

Ejemplo

En la tabla que viene a continuación se presentan los estadísticos iniciales obtenidos en nuestro ejemplo. El valor propio λ_k indica la cantidad de varianza restituida por el factor k. Cuanto más valor tiene λ_k, más es alto el nivel explicativo del factor asociado.

Variable	Comunalidad	*	Factor	Valor propio λk	Varianza explicada	Varianza acumulada
V1	1.00000	*	1	10.31684	51.6	51.6
V2	1.00000	*	2	5.74034	28.7	80.3
V3	1.00000	*	3	1.39340	7.0	87.3
V4	1.00000	*	4	1.02887	5.1	92.8
V5	1.00000	*	5	.72643	3.6	96.0
V6	1.00000	*	6	.41184	2.1	98.1
V7	1.00000	*	7	.16824	.8	98.9
V8	1.00000	*	8	.14197	.7	99.6
V9	1.00000	*	9	.05583	.3	99.9
V10	1.00000	*	10	.01623	.1	100.0
V11	1.00000	*	11	.00000	.0	100.0
V12	1.00000	*	12	.00000	.0	100.0
V13	1.00000	*	13	.00000	.0	100.0
V14	1.00000	*	14	.00000	.0	100.0
V15	1.00000	*	15	.00000	.0	100.0
V16	1.00000	*	16	.00000	.0	100.0
V17	1.00000	*	17	.00000	.0	100.0
V18	1.00000	*	18	.00000	.0	100.0
V19	1.00000	*	19	.00000	.0	100.0
V20	1.00000	*	20	.00000	.0	100.0

Estadísticos iniciales.

A partir de estos valores podemos calcular el porcentaje de varianza restituida por cada factor aplicando la fórmula siguiente:

$$\%\ V(F_k) = \frac{\lambda_k}{\sum_{k=1}^{K} \lambda_k}.$$

Ejemplo

En nuestro ejemplo, el primer factor resume el 51,6% de la varianza total; el segundo, el 28,7%; el tercero, el 7%, y el décimo factor, el 0,1%. Con los dos primeros factores el análisis restituye el 80,3% de la información contenida en la matriz de correlaciones expuesta.

En cuanto a la comunalidad, indica la proporción de varianza de cada variable explicada por los factores seleccionados. Al trabajar inicialmente con todos los factores, la comunalidad de cada variable es máxima, es decir, igual a 1.

Profundizaremos en el aspecto de la comunalidad un poco más adelante.

Etapa 4: determinación del número de factores que hay que conservar

En nuestro ejemplo,...

... seleccionaríamos los cuatro primeros factores.

Hay una gran cantidad de reglas y criterios para determinar cuál es el número de ejes factoriales que hay que conservar. La mayor parte de los programas de ordenador más conocidos suele aplicar el denominado **criterio de Kaiser**, según el cual sólo se conservan los factores cuyos valores propios, λ_k, son más altos que la unidad. Los criterios más utilizados pueden agruparse en dos métodos generales:

1) Reglas basadas en la restitución mínima

En este caso, lo que hace el investigador es fijar *a priori* un nivel correspondiente al porcentaje mínimo de varianza que quiere restituir y conservar para el análisis el número de ejes necesario para alcanzar este nivel.

Por ejemplo, sabemos que el porcentaje de varianza explicada por los dos primeros factores es el siguiente:

$$\%\ V(F_1,F_2) = \frac{\lambda_1 + \lambda_2}{\sum_{k=1}^{K} \lambda_k}.$$

En el caso de que esta cantidad alcance el nivel fijado, tenemos que conservar únicamente estos dos primeros factores; en el caso contrario, introduciremos en el análisis el tercer factor y así sucesivamente hasta alcanzar el nivel fijado.

Por ejemplo, supongamos que el nivel fijado sea, en porcentaje, el 90%; es preciso conservar, pues, para el análisis el número de factores que haga que:

$$\% V(F_1,\ F_2,\ldots,\ F_k)\ =\ \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\sum\limits_{k=1}^{K} \lambda_k} \geq 90\%\,.$$

En nuestro ejemplo,...

... seleccionaríamos los cuatro primeros factores.

2) Reglas basadas en la información restituida por cada factor

Las tres reglas que se exponen a continuación, basadas en la cantidad de información restituida por cada factor, son reglas empíricas obtenidas después de numerosos análisis; no tienen, como es lógico, ninguna demostración teórica, pero se basan en el sentido común.

a) Primera regla empírica. Únicamente es preciso conservar para el análisis aquellos factores que restituyan una proporción de la varianza superior a dos veces la cantidad 100/*P*; *P* es el número de variables iniciales introducidas en el análisis.

Esta regla proviene del hecho de que si la nube de puntos no tiene ninguna dirección privilegiada (esfera, por ejemplo), los valores difieren muy poco y la varianza restituida por el primer factor sería, más o menos:

$$V(F_1)\ =\ \frac{\lambda_1}{\sum\limits_{k=1}^{K} \lambda_k}\ =\ \frac{1}{P'}$$

o bien el porcentaje 100/*P*, donde *P* es el número de variables iniciales; después, todos los valores propios serían aproximadamente iguales.

Ejemplo

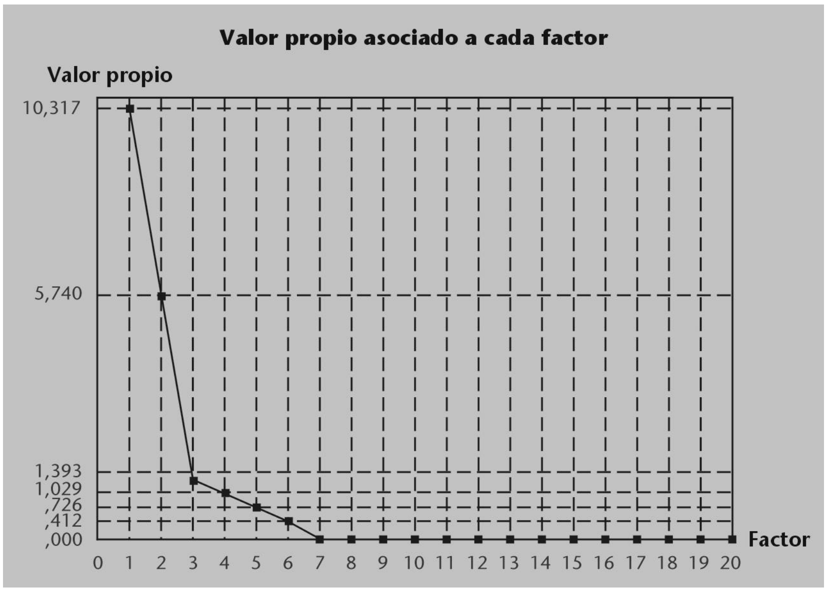
En nuestro ejemplo seleccionaríamos los dos primeros factores:

$$2 \cdot \frac{100}{20}\ =\ 10$$

b) Segunda regla empírica. Se trata de construir una curva en la cual los puntos sean los siguientes:

- en abscisas, los números de los factores;
- en ordenadas, el porcentaje de varianza que restituye cada uno de los factores o el valor propio asociado a cada factor.

Hay que determinar el primer punto de inflexión de la curva y conservar aquellos factores cuyo número de orden esté situado antes del punto de inflexión, tal como se expone en el gráfico que viene a continuación. En nuestro ejemplo, el cambio de concavidad se produce a partir del tercer factor. Conservaríamos, por tanto, los dos primeros factores.



c) Tercera regla empírica. También se denomina *regla de interpretación*, y es la regla más utilizada en investigación de mercados, ya que tiene en cuenta la facilidad de interpretación y la operatividad de los factores extraídos. Selecciona el número de factores necesarios para cumplir los dos criterios siguientes:

- La solución debe ser fácilmente interpretable, es decir, tiene que comunicar de forma tan fiel como sea posible la configuración inicial de variables.
- Los factores tienen que ser operativos, es decir, de fácil utilización como variables relevo en estudios o análisis posteriores.

Ejemplo

En nuestro ejemplo, decidimos inicialmente conservar los dos primeros factores, con lo que conservamos el 80,3% de la información inicial, tal como se expone en la tabla siguiente:

Variable	Comunalidad	*	Factor	Valor propio λk	Varianza explicada	Varianza acumulada
V1	.88066	*	1	10.31684	51.6	51.6
V2	.69564	*	2	5.74034	28.7	80.3
V3	.94876	*	3			
V4	.92548	*	4			
V5	.85292	*	5			
V6	.83597	*	6			
V7	.73753	*	7			
V8	.37139	*	8			
V9	.42255	*	9			
V10	.67027	*	10			
V11	.97686	*	11			

Variable	Comunalidad	*	Factor	Valor propio λk	Varianza explicada	Varianza acumulada
V12	.35722	*	12			
V13	.88646	*	13			
V14	.96759	*	14			
V15	.98637	*	15			
V16	.77473	*	16			
V17	.95604	*	17			
V18	.95938	*	18			
V19	.93150	*	19			
V20	.91990	*	20			

Solución con los dos primeros factores.

Evidentemente, la decisión final depende de la facilidad de interpretación de los factores y de la calidad de la información conservada.

Un primer resultado que indica la calidad de la representación de la información en los dos factores seleccionados es la **comunalidad**. La comunalidad expresa la proporción de varianza de una variable explicada por los factores seleccionados. Concretamente, indica el porcentaje de información de cada variable que estamos perdiendo al trabajar en un espacio determinado.

Una comunalidad elevada (próxima a 1) implica una correlación elevada con **al menos uno** de los factores seleccionados; en cambio, una comunalidad baja implica una correlación baja con **todos** los factores seleccionados. Estas variables están correlacionadas con otros factores.

Ejemplo

En nuestro ejemplo, teniendo en cuenta los dos primeros factores, la mayoría de las variables está bien representada, tal como podéis ver en el gráfico anterior, a excepción de las siguientes:

	Comunalidad
V8 Visten con gusto, van limpios y aseados	0.37139
V9 Son puntuales	0.42255
V10 Son tranquilos, tienen control de sí mismos	0.67027
V12 Son ordenados en sus explicaciones	0.35722

Etapa 5: rotación de los factores conservados

Como hemos expuesto más arriba, los ejes factoriales pueden considerarse las “dimensiones latentes” del problema, y describirlas (interpretarlas) nos conduce a comprender las dimensiones fundamentales del fenómeno que es objeto de estudio.

Para interpretar de forma correctamente estas dimensiones, necesitamos saber cuáles son las variables que contribuyen más a la formación de cada factor. Con esta finalidad, utilizaremos como indicador los **coeficientes de correlación entre las variables iniciales y los factores** que en la fase anterior hayamos decidido conservar para proseguir el análisis. Las variables con los coeficientes de correlación más altos con un factor son las que contribuyen más a la formación de este factor.

La matriz que contiene los coeficientes de correlación entre las variables iniciales y los factores se suele denominar *matriz factorial inicial* o *matriz factorial no rotada*.

	Factor 1	Factor 2
V13	.90241	.26855
V19	.86195	-.43421
V6	.85212	-.33144
V2	.83227	.05452
V15	.82224	-.55703
V10	.81793	.03563
V7	.81424	-.27302
V16	.79696	-.37360
V11	.79174	-.59162
V3	.77722	.58711
V14	.76988	-.61226
V20	.75042	-.59730
V4	.72068	.63725
V8	.55348	.25504
V12	.54456	-.24632
V17	.52147	.82711
V18	.53442	.82084
V1	.46372	.81585
V5	.62024	.68427
V9	.17677	.62554

Ejemplo de matriz factorial no rotada.

Aunque esta matriz indica las relaciones entre los factores y las variables iniciales, raramente estas últimas pueden interpretarse con facilidad, ya que suele suceder que algunas variables iniciales están altamente correlacionadas con varios factores.

Ejemplo

En nuestro ejemplo, las variables V4 y V14 tienen correlaciones bastante elevadas en ambos factores.

Para solucionar este problema, suele efectuarse lo que se denomina una **rotación de los factores**, que consiste en transformar la matriz factorial inicial en una matriz factorial rotada de interpretación más fácil. Se trata de que cada factor tenga coeficientes de correlación significativos con tan sólo algunas de las variables iniciales, y que cada variable inicial tenga coeficientes de correlación significativos con tan sólo algunos factores, si es posible sólo con uno.

Hay dos tipos de rotaciones:

1) Las rotaciones oblicuas, que son las que eliminan la propiedad de independencia de los factores.

2) Las rotaciones ortogonales, que son las que la mantienen.

En investigación comercial, suelen aplicarse las rotaciones ortogonales, dada la complejidad en la interpretación de resultados de las rotaciones oblicuas.

Dentro de las rotaciones ortogonales, uno de los algoritmos más utilizados es el algoritmo VARIMAX. Este algoritmo intenta, para cada factor, maximizar la correlación de algunas variables, las más próximas a –1 o +1, y minimizar la correlación del resto de las variables.

Ejemplo

En nuestro ejemplo, si aplicamos una rotación VARIMAX a la matriz factorial no rotada, obtenemos el resultado expuesto a continuación. Efectivamente, observamos que las variables con coeficientes de correlación próximos a 1 en el factor 1 presentan coeficientes de correlación próximos a 0 en el factor 2, y las variables con coeficientes de correlación próximos a 1 en el factor 2 presentan coeficientes de correlación próximos a 0 en el factor 1.

	Factor 1	Factor 2
V15	.99312	-.00846
V11	.98688	-.05415
V14	.98011	-.08344
V19	.95817	.11580
V20	.95562	-.08176
V6	.89307	.19594
V16	.87049	.13028
V7	.82017	.22360
V2	.66279	.50631
V10	.66132	.48263
V12	.58984	.09648
V18	-.00960	.97943
V17	-.02386	.97748
V1	-.06570	.93613
V4	.24717	.92972
V3	.32201	.91928
V5	.13749	.91325
V13	.60267	.72336
V9	-.19923	.61875
V8	.31961	.51888

Ejemplo de matriz factorial rotada.

Al efectuar una rotación, hay que tener en cuenta que el total de la información restituida (en nuestro ejemplo, el 80%) permanece constante, pero varía la información restituida por cada uno de los factores; por eso, si hay que conocerla, tiene que recalcularse. Si llamamos b_{pk} al peso de la variable p en el factor k rotado, el porcentaje de varianza explicada por este factor es:

$$\% V(F_k) = \frac{\sum_{p=1}^{P} b_{pk}^2}{\sum_{p=1}^{P} \sum_{k=1}^{K} b_{pk}^2},$$

donde P es el número de variables, y K, el número de factores que contiene la matriz factorial.

Ejemplo

En nuestro ejemplo, la varianza explicada por cada uno de los factores, antes y después de efectuar la rotación, es:

	Antes de rotar	Después de rotar
Factor 1	51,6	44,6
Factor 2	28,7	35,7
Factor 1 + 2	80,3	80,3

Etapa 6: interpretación de la matriz factorial rotada y representación de los resultados

El objetivo de la interpretación de la matriz factorial rotada consiste en identificar cada una de las dimensiones latentes extraídas. Se efectúa eligiendo para cada factor las variables iniciales que tengan unas correlaciones con el factor que sean las más elevadas (próximas a +1 ó a –1).

Ejemplo

En nuestro caso, para cada factor las variables más correlacionadas son:

	Factor 1	
V15	Respuestas ingeniosas y completas	0.99312
V11	Parece seguro de sí mismo	0.98688
V14	Docto en la materia. Desarrolla la asignatura con sentido crítico	0.98011
V19	Buena dicción, plenamente audible	0.95817
V20	Sabe controlar la situación con facilidad, sin nerviosismo	0.95562
	Factor 2	
V18	Sabe apreciar los esfuerzos realizados por el estudiante	0.97943
V17	Es tolerante con los errores de los estudiantes	0.97748
V1	Trata a todos los estudiantes aproximadamente igual	0.93613
V4	Muestra interés por los problemas y las necesidades de los estudiantes	0.92972
V3	Próximo, cercano a los estudiantes	0.91928
V5	Posee un carácter agradable	0.91325

A la vista de las variables que constituyen cada uno de los factores, vemos que el factor 1 tiene relación con aspectos referentes a la calidad docente de los profesores, y el factor 2, con aspectos referentes a la calidad humana de los profesores. De esta manera, podríamos bautizar el factor 1 como “buen profesor” y el factor 2 como “buena persona”.

4.1.3. Aplicaciones del análisis factorial

Las principales aplicaciones del análisis factorial son las siguientes:

1) **Utilización de los resultados del análisis factorial de componentes principales como “variables-relevo”**. Los resultados de un análisis factorial pueden utilizarse como fase previa de cálculo antes de aplicar otros métodos. Por ejemplo, dado que los factores obtenidos son independientes, podemos utilizarlos como variables nuevas, y evitar dificultades en el caso de que haya una correlación estrecha entre las variables iniciales, que es uno de los problemas más comunes en análisis como regresión múltiple, análisis tipológico o análisis discriminante.

2) **Selección de variables.** Este tipo de análisis permite seleccionar, de entre un conjunto importante de variables, cuáles son las que más intervienen en la descripción del fenómeno estudiado, y posibilita que conservemos para análisis posteriores únicamente aquellas variables iniciales que estén estrechamente correlacionadas con los factores que hemos considerado más importantes.

3) Detección de conglomerados. Si efectuamos un análisis factorial de tipo Q, que recordamos que consiste en buscar los factores en el espacio de los individuos, podemos agrupar a los individuos en función de comportamientos análogos en relación con las variables sobre las cuales se lleva a término el análisis, utilizando por ejemplo un algoritmo de clasificación no jerárquico.

Podéis consultar el subapartado 4.3 de este módulo para el estudio del análisis tipológico no jerárquico.

4.2. Análisis de correspondencias

4.2.1. Cuándo tenemos que utilizar el análisis de correspondencias

Hay varias maneras de presentar el análisis de correspondencias (ACOR). Quizá la más correcta y comprensible sea decir que este tipo de análisis es un caso particular del análisis factorial de componentes principales (ACP), y se diferencia básicamente en el aspecto siguiente: mientras que un ACP trata de resumir el número de variables que intervienen en un análisis por medio de la construcción de nuevas variables compuestas (o factores) más sintéticas, en un ACOR se trata de analizar las formas que adoptan las relaciones entre las categorías de las variables.

Ejemplo de análisis factorial de correspondencias

Supongamos que la información de la que disponemos es la que se expone a continuación:

En un estudio sobre el mercado de material eléctrico se entrevistó a una muestra de 1.318 clientes de nueve empresas de distribución, con el fin de determinar el posicionamiento de estas empresas respecto de siete atributos que habían sido definidos previamente como de una gran importancia dentro de su sector de actividad. Cada cliente tuvo que asociar a cada empresa uno de los siete atributos siguientes:

- Ofrece los mejores precios o descuentos.
- Tiene más variedad de marcas.
- Ofrece más rapidez de entrega.
- Proporciona una mejor información técnica o consejos.
- Ofrece un mejor trato del personal.
- Ofrece unas mejores condiciones de pago.
- Es más fácil de acceder a ésta.

Los resultados del estudio se presentan en forma de una tabla de contingencia o correspondencias:

	Emp. 1	Emp. 2	Emp. 3	Emp. 4	Emp. 5	Emp. 6	Emp. 7	Emp. 8	Emp. 9	Totales fila
Precios o descuentos	16 (*)	17	18	19	16	45	15	19	18	183
Variedad de marcas	8	15	18	17	27	20	2	14	53	174
Rapidez de entrega	20	20	23	21	29	20	18	19	25	195
Información técnica	11	13	12	17	20	16	15	10	44	158
Trato del personal	8	25	25	22	30	26	24	22	26	228
Condiciones de pago	21	21	20	24	27	22	18	21	24	198
Facilidad de acceso	21	21	21	23	26	15	16	18	21	182
Totales columna	125	132	137	143	175	164	108	123	211	1.318

(*) 16 clientes han asociado el atributo "mejores precios o descuentos" a la empresa 1.