INVESTIGACIÓN en Cómic

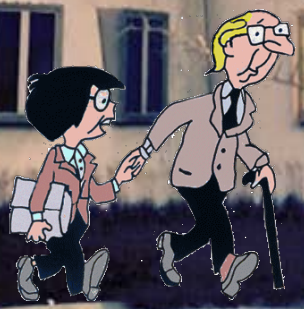
Análisis Multivariante

Análisis Factorial

Componentes Principales (ACP)

Análisis Cluster

SPSS



Me tienes que perdonar sino consigo atraer tu atención.

SPSS

El *análisis factorial* es un método que intenta explicar (según un modelo lineal) un conjunto grande de variables observadas mediante un número reducido de variables hipotéticas llamadas factores.

Matemáticamente genera varias combinaciones lineales de las variables observadas. Cada combinación lineal forma un factor.

La respuesta a una variable es combinación lineal de factores comunes, factores específicos y factores de error:

$$X_i = F_{i1} \times A_{i1} + F_{i2} \times A_{i2} + \ldots\ldots + F_{in} \times A_{in} + A_{ij} \times E_i$$

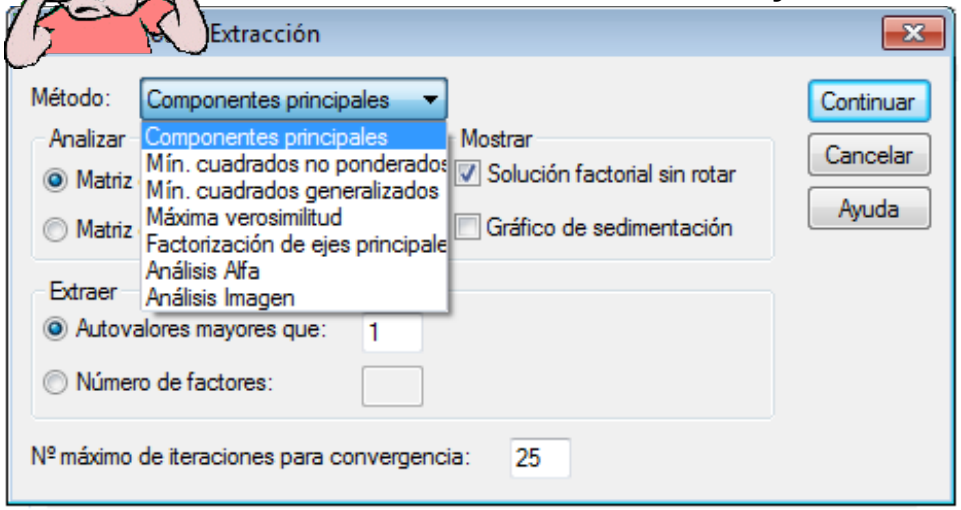
La variable X_i queda influida por los factores, el influjo propio de la variable (especificidad) y el error.

X_i = puntuación de un individuo en la variable i-ésima
F_{i1} = coeficiente del sujeto en el factor común 1
A_{i1} = peso factorial de la variable i en el factor común 1
E_i = puntuación del sujeto en el factor error

Tipos de Análisis Factorial

- Factorización Ejes Principales
- Componentes Principales
- Mínimos cuadrados no ponderados
- Mínimos cuadrados generalizados
- Máxima verosimilitud
- Análisis Alfa
- Análisis imagen

Los tipos de análisis factorial se definen considerando el proceso de extracción de los factores y el método de cálculo de las comunalidades.



El análisis de componentes principales fue desarrollado por Harold Hotelling en 1933. Actualmente es el método más utilizado en investigación social y comercial.

El método presupone que no hay factores comunes. Por tanto, más que reducir el número de variables trata de simplificar la estructura de los datos transformando las variables en unas pocas componentes principales que sean combinaciones lineales de las variables.

En otras palabras, esta técnica estadística analizando la estructura de dependencia y correlación que existe entre las variables, trata de transformar un conjunto de *variables interrelacionadas* en un conjunto de *variables no correlacionadas* llamadas *factores*, especificando como éstas explican parte de la información que contienen los factores.

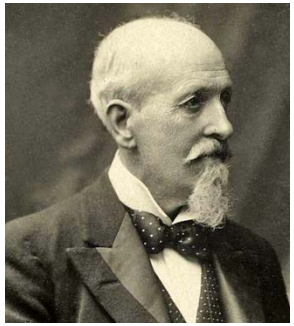
Tratando de ajustar de forma optima una nube de puntos sobre un espacio de 1, 2, … , n dimensiones

El método establece que las comunalidades valen 1

- Se elige primero el factor (factor 1) que explique la mayor parte de la varianza de las variables.
Este primer componente puede considerarse como el resumen óptimo de las relaciones lineales existentes.

- Se resta el peso del primer factor a las variables. Con la variabilidad restante (residual) se calcula la segunda componente principal (factor 2), y así sucesivamente.

En resumen, las componentes principales son una combinación lineal de las variables observadas.
El producto final es una matriz transformada: Las columnas son los vectores propios de la matriz de correlaciones.



La *Factorización de Ejes Principales*, también conocido como Factores Principales o Factor Común, fue introducido por Charles Edward Spearman.

Este método es utilizado frecuentemente por psicólogos.

La técnica estadística considera que dos variables, por lo que tienen en común, están relacionadas por su comunalidad. De modo, que existen una serie de correlaciones comunes que se tendrán que localizar.

Cada variable, además del componente común, tiene una parte específica que no está relacionada con las demás, siendo explicada por factores específicos o únicos.

Cada variable, además del componente común, tiene una parte específica que no está relacionada con las demás, siendo explicada por factores específicos o únicos.

Dentro de cada variable se diferencia la parte común y otra específica o única. De esta forma, un factor se interpreta por lo que tienen en común las variables con las que está relacionado.

La forma de hacer es parecida al método de componentes principales. En la *Factorización de Ejes Principales*, la diagonal de la matriz de correlaciones está formada por las estimaciones de la comunalidad.

- Se utilizan los coeficientes de correlación múltiple al cuadrado como estimadores iniciales de la comunalidad, y desde aquí se extrae el número necesario de factores.
- Las comunalidades se vuelven a estimar a partir de los pesos factoriales. Los factores se extraen de nuevo con las nuevas comunalidades que reemplazan a las anteriores. El proceso iterativo continúa hasta que no hay un cambio significativo en las comunalidades estimadas.

Diferencias entre Componentes Principales y Factorización de Ejes Principales

La principal diferencia está en la varianza que es analizada.

El análisis de componentes principales analiza la totalidad de la varianza, mientras que el análisis de ejes principales analiza la covarianza (comunalidad).

Análisis componentes principales

Cada variable en la explicación de la varianza total al principio aporta un coeficiente igual a 1.

El objetivo es extraer la máxima varianza a partir de los datos, identificando a unos pocos componentes ortogonales.

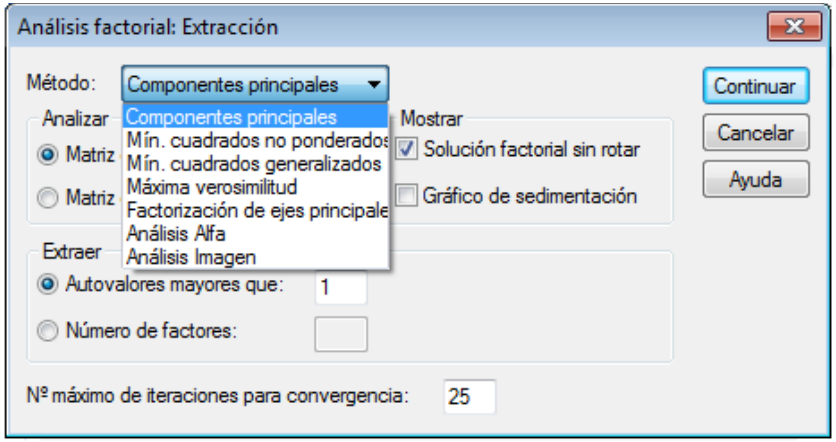
Cuando interesa un resumen empírico de los datos.

Análisis ejes principales

Se centra en las variables con una alta comunalidad.

El objetivo es reproducir la matriz de correlaciones con un número de componentes.

Cuando a partir de un modelo teórico se quiere inferir soluciones hipotéticas.



TIPOS DE ANÁLISIS FACTORIAL

Mínimos cuadrados no ponderados: Genera una matriz de pesos factoriales que minimiza las sumas de los cuadrados de las diferencias entre las matrices de correlaciones observada y reproducida, ignorando los elementos de la diagonal.

Mínimos cuadrados generalizados: Con un criterio igual al anterior, pondera los los coeficientes de correlación inversamente a la unicidad de las variables. Las variables con alta unicidad (baja comunalidad) tienen una influencia pequeña en el resultado final. Genera un estadístico de bondad de ajuste *Chi-cuadrado* que contrasta la hipótesis nula de que la matriz residual es una matriz nula.

Máxima verosimilitud: Calcula las estimaciones de los parámetros que con mayor probabilidad han producido la matriz de correlaciones observada, asumiendo que la muestra procede de una distribución normal multivariada. Es decir, busca la solución factorial que mejor se ajusta a las correlaciones observadas. Pondera los coeficientes inversamente a los valores de la unicidad, utilizando un algoritmo iterativo para su cálculo.
Genera un estadístico de bondad de ajuste *Chi-cuadrado* que contrasta el grado de ajuste entre lo real y lo estimado, con lo que permite determinar el número de factores necesarios para lograr el mejor ajuste.

TIPOS DE ANÁLISIS FACTORIAL

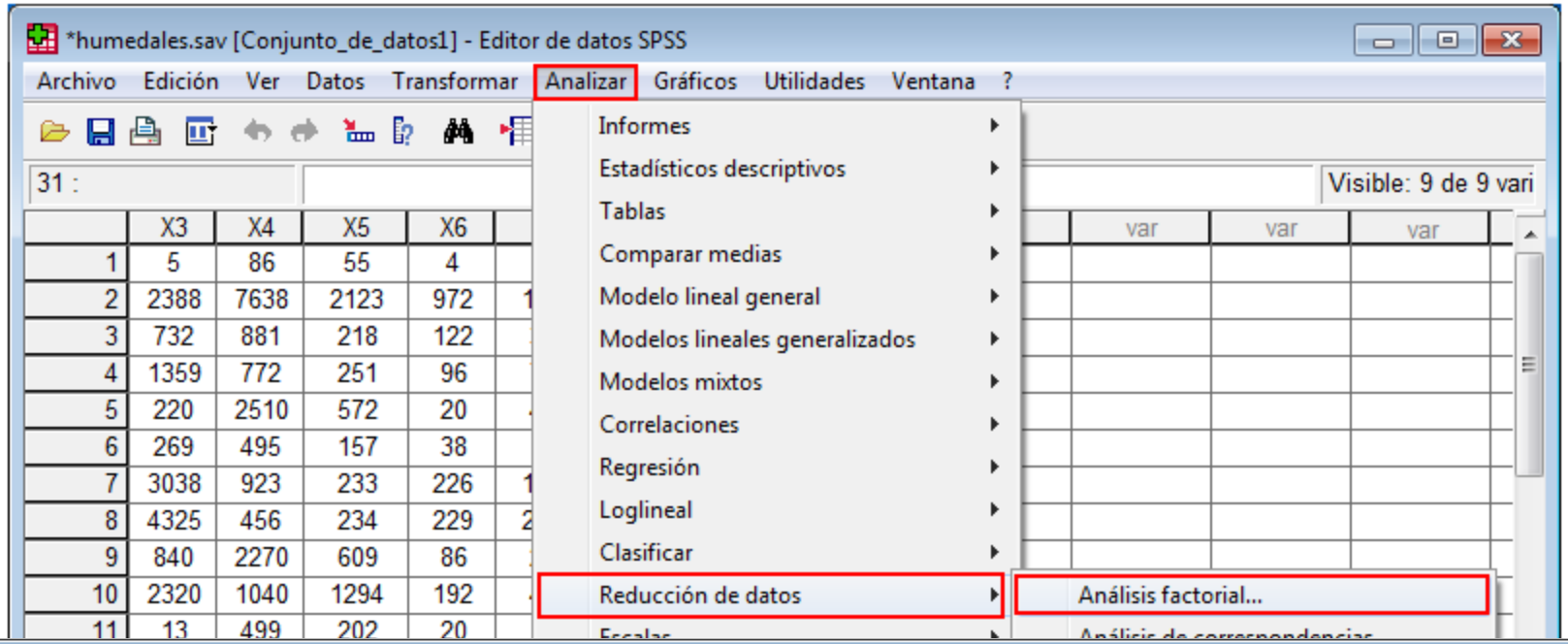
Análisis Alfa: Considera las variables incluidas en el análisis como una muestra del conjunto posible de variables, intentando maximizar la fiabilidad de los factores respecto a la totalidad de las variables. Para ello, el método extrae una serie de factores comunes que tengan correlaciones máximas con el conjunto de factores comunes existentes, mientras que los factores específicos son los generados por el muestreo aleatorio.

Análisis Imagen: Mediante la correlación múltiple, estudia las partes comunes y únicas de las variables observadas.
La imagen de una variables es la parte que tiene en común la variable con otras variables, a diferencia de la anti-imagen que es la parte exclusiva de cada variable.

Los datos que se introducen en SPSS corresponden a la medición de 17 humedales en determinada época del año.

X1: Conductividad eléctrica
X2: Contenido en bicarbonatos
X3: Contenido en cloruros
X4: Contenido en sulfatos
X5: Contenido en calcio
X6: Contenido en magnesio
X7: Contenido en sodio
X8: Contenido en potasio
X9: Contenido en fosfatos

Analizar/Reducción de datos/Análisis factorial...



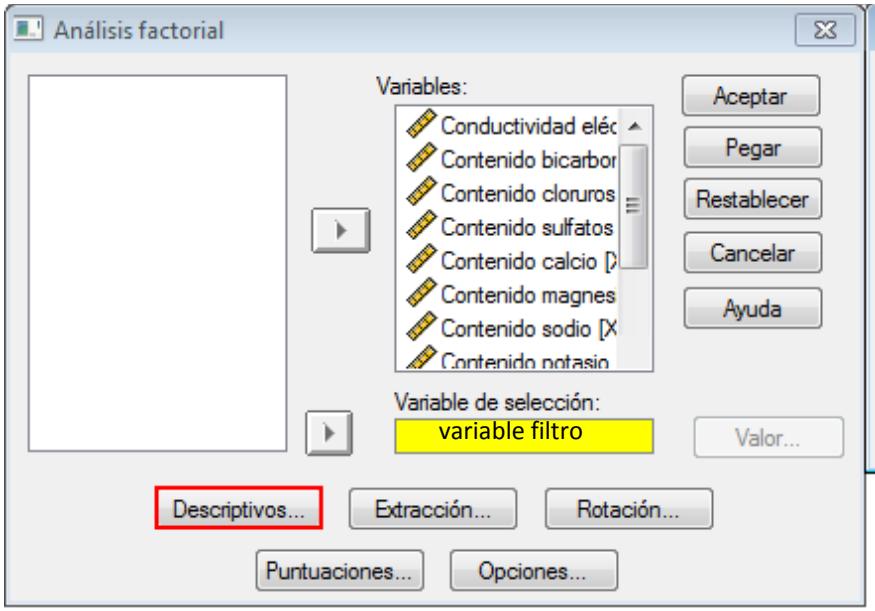
*humedales.sav [Conjunto_de_datos1] - Editor de datos SPSS

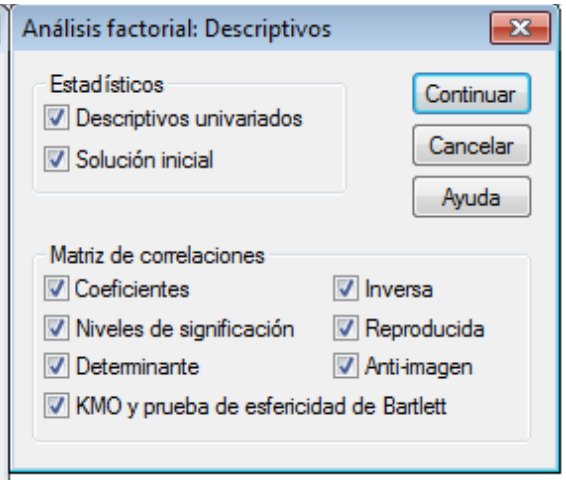
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	X1	Numérico	8	2	Conductividad eléctrica	Ninguno	Ninguno	5	Centrado	Escala
2	X2	Numérico	8	4	Contenido bicarbonatos	Ninguno	Ninguno	6	Centrado	Escala
3	X3	Numérico	8	0	Contenido cloruros	Ninguno	Ninguno	5	Centrado	Escala
4	X4	Numérico	8	0	Contenido sulfatos	Ninguno	Ninguno	5	Centrado	Escala
5	X5	Numérico	8	0	Contenido calcio	Ninguno	Ninguno	6	Centrado	Escala
6	X6	Numérico	8	0	Contenido magnesio	Ninguno	Ninguno	5	Centrado	Escala
7	X7	Numérico	8	0	Contenido sodio	Ninguno	Ninguno	6	Centrado	Escala
8	X8	Numérico	8	0	Contenido potasio	Ninguno	Ninguno	5	Centrado	Escala
9	X9	Numérico	8	4	Contenido fosfatos	Ninguno	Ninguno	6	Centrado	Escala
10										

Vista de datos | Vista de variables

SPSS El procesador está preparado





Variable filtro para definir una sub-muestra de sujetos que cumplen una condición.

- Descriptivos univariados: Para cada variable muestra el número de casos válidos, la media y la desviación típica.

- Solución inicial: Permite obtener las comunalidades iniciales, autovalores de la matriz analizada y los porcentajes de varianza asociados a cada autovalor. Está opción actúa por defecto y la información que ofrece aparece en las tablas de Comunalidades y Porcentaje de varianza explicada.

- Coeficientes: Muestra la matriz con los coeficientes de correlación entre las variables del análisis.

- Niveles de significación: En la matriz de correlaciones incluye los niveles críticos unilaterales asociados a cada coeficiente.

- Determinante: Muestra el determinante de la matriz de correlaciones. Los determinantes próximos a 0 indican que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial es una técnica adecuada para analizar estas variables.

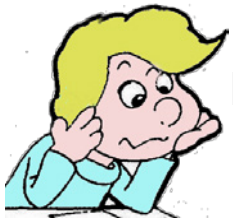
- Inversa: Muestra la inversa de la matriz de correlaciones, base para el cálculo de las comunalidades iniciales en algunos modelos de extracción y para el cálculo de la matriz anti-imagen.

- Reproducida: La matriz reproducida es la que se obtiene a partir de la solución factorial hallada. Si el modelo es bueno y el número de factores el adecuado, la estructura factorial debe poder reproducir la matriz de correlaciones.
En la diagonal de la matriz reproducida se encuentran las comunalidades finales.
Junto con la matriz de correlaciones reproducidas aparece la matriz de correlaciones residuales, que contiene los residuos, esto es, las diferencias entre las correlaciones observadas y las correlaciones reproducidas.
Cuando el modelo es el correcto, debe ser mínimo el número de residuos con valores elevados.

- Anti-imagen: Muestra la matriz de covarianzas anti-imagen y la matriz de correlaciones anti-imagen. La matriz de covarianzas anti-imagen contiene los negativos de las covarianzas parciales. La matriz de correlaciones anti-imagen contiene los coeficientes de correlación parcial cambiados de signo (la correlación entre dos variables se parcializa considerando el resto de variables del análisis).

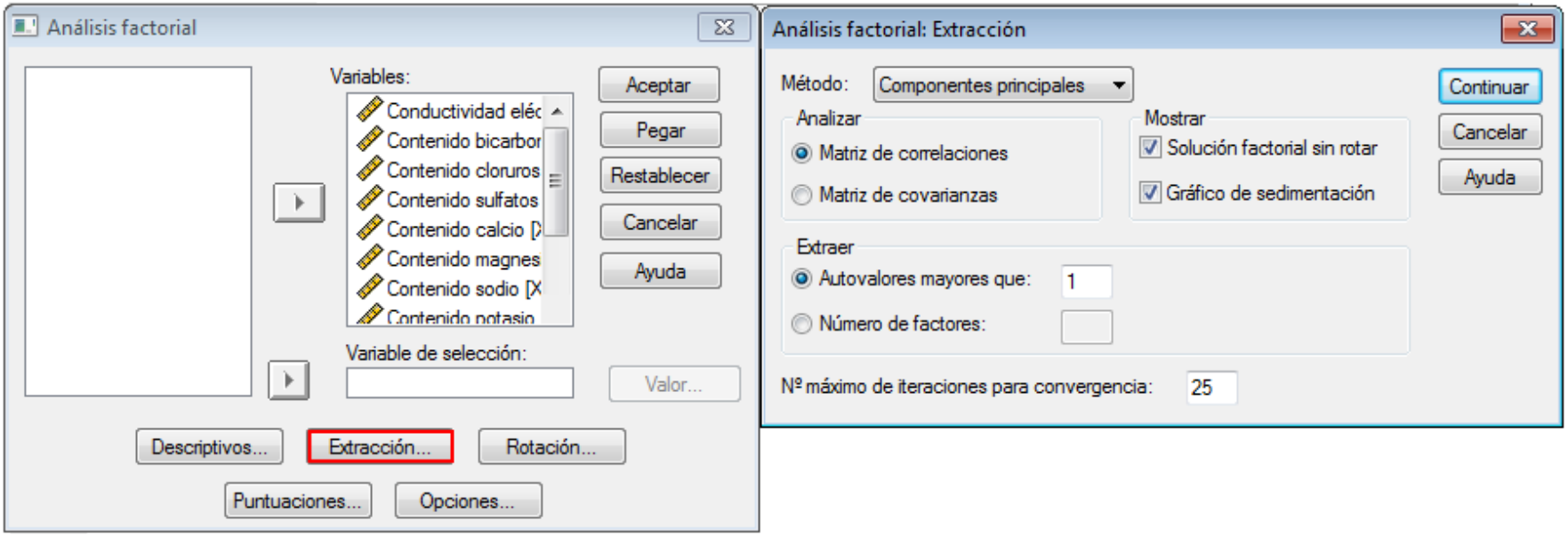
• Anti-imagen: Si el modelo factorial elegido es adecuado para explicar los datos, los elementos de la diagonal de la matriz de correlaciones anti-imagen deben de tener un valor próximo a 1 y el resto de los elementos deben de ser pequeños.

• KMO y prueba de esfericidad de Bartlett: La medida de Kaiser-Meyer-Olkin contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas. Permite comparar la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial.
El estadístico KMO varía entre 0 y 1. Los valores pequeños anuncian que el análisis factorial puede no ser adecuado, dado que las correlaciones entre los pares de variables no pueden ser explicados por otras variables.
Los valores de KMO menores que 0,5 indican que no debe utilizarse el análisis factorial con los datos muestrales que se están utilizando en el análisis.
📖 La prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las variables y el modelo factorial no sería adecuado.



En el *Visor* de SPSS...

Extracción de factores



- Autovalores mayor que: Cuando se analiza la matriz de correlaciones esta opción permite utilizar el tamaño de los autovalores como criterio para decidir el número de factores que se encuentran en la solución factorial.
Cuando la matriz es la de covarianzas, la opción decide el número de veces que un autovalor debe de ser mayor que el autovalor promedio de la matriz para que el correspondiente factor sea retenido en la solución.

- Número de factores: Se puede especificar el número exacto de factores que se desea incluir en la solución.

- Solución factorial sin rotar: Muestra las saturaciones sin rotar (matriz de componentes o factorial), comunalidades y autovalores de la solución factorial.

• Gráfico de sedimentación: Es una representación gráfica de la magnitud de los autovalores. En tendencia descendente, el corte se utiliza para determinar el número óptimo de factores que deben estar presentes en la solución. Independientemente del método de extracción seleccionado, siempre se muestra la representación de los autovalores de la matriz de correlaciones (o covarianzas) originales.

• Nº de iteraciones para convergencia: Se puede especificar el número máximo de iteraciones que los algoritmos pueden realizar para encontrar la solución factorial final.

En el Visor de SPSS aparece el cálculo de los 9 factores extraídos, sus autovalores, el porcentaje de varianza explicado por cada uno y el porcentaje acumulado.

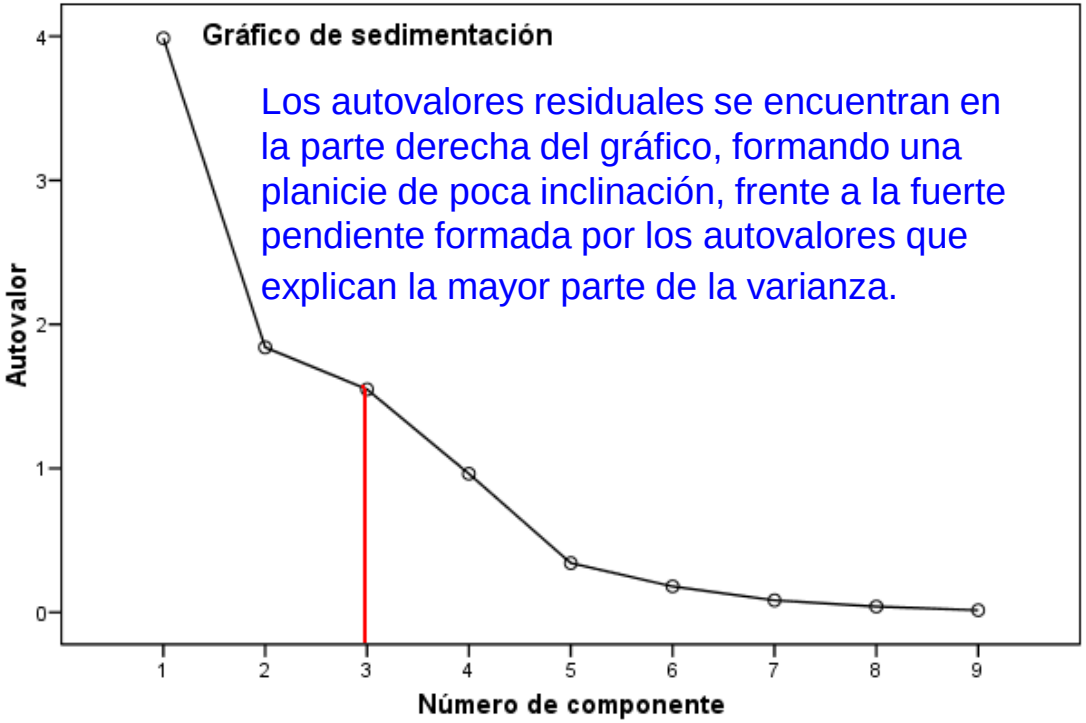
Como el objetivo del análisis factorial es reducir las 9 variables a un número menor de dimensiones, el investigador debe decidir el número de factores con los que desea trabajar.

El investigador puede seguir varias estrategias para conseguir el objetivo de tener un menor número de dimensiones que expliquen la mayor parte de la varianza.

📖 Con la opción *Extraer Autovalores mayores que* puede seleccionar mayores que un determinado tamaño, en la medida que autovalores mayores explican un mayor porcentaje de varianza.

📖 Fijar un porcentaje de varianza explicada y tomar los factores que lleguen a explicarlo.

📖 Mediante una solución gráfica (*Gráfico de sedimentación*) ir seleccionando los factores hasta obtener un cambio de tendencia en la línea que une estos puntos, una ruptura de la continuidad que se manifiesta en un alisamiento de la pendiente.
Un cambio de tendencia indica un dominio de la varianza única respecto a la varianza común.



Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,986	44,291	44,291	3,986	44,291	44,291
2	1,840	20,448	64,739	1,840	20,448	64,739
3	1,549	17,207	81,946	1,549	17,207	81,946
4	,962	10,693	92,639			
5	,343	3,807	96,446			
6	,180	2,005	98,451			
7	,084	,933	99,384			
8	,041	,452	99,836			
9	,015	,164	100,000			

Método de extracción: Análisis de Componentes principales.

Para calcular el porcentaje de varianza explicado con cada factor, hay que considerar que con los 9 factores se explica la varianza en su totalidad. Por tanto, la suma de todos los autovalores será 9.

A continuación se considera la magnitud de cada autovalor en relación al porcentaje total de varianza explicada: La suma de los 9 autovalores explica el 100% de la varianza.

- El autovalor con magnitud 3,986 (primer factor) explica el 44,291% de la varianza (3,986/9).
- El segundo autovalor, con magnitud 1,840, explica el 20,448% de la varianza (1,840/9).
- Así sucesivamente.

De este modo, con los 3 primeros autovalores se explica el 81,946% de la varianza. A medida que se aumentan los autovalores, aumenta el ajuste entre los datos observados y el modelo construido, disminuyendo la parsimonia del modelo y complicando la interpretación de los ejes.

Primera evaluación …

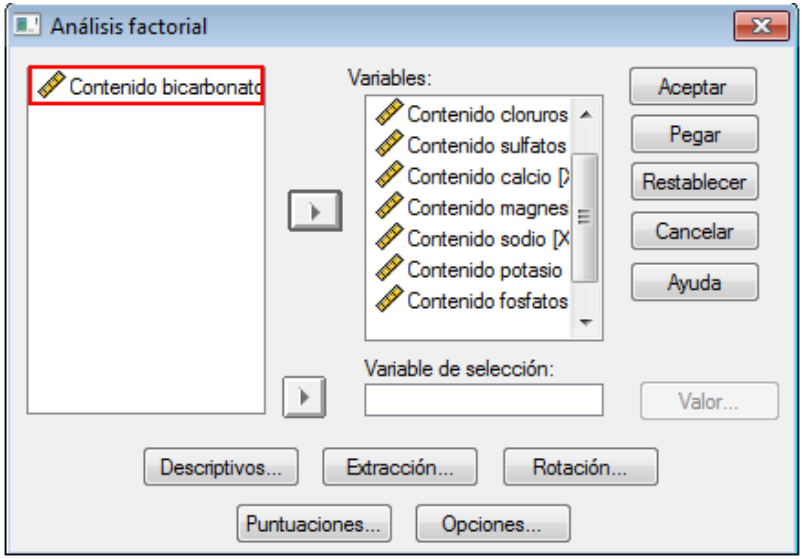
El investigador, al observar la matriz de comunalidades, puede ir ajustando el modelo eliminando las variables que tienen una extracción menor o cercana a 0,32 (evaluación pobre).

Comunalidades

	Inicial	Extracción
Conductividad eléctrica	1,000	,936
Contenido bicarbonatos	1,000	,338
Contenido cloruros	1,000	,946
Contenido sulfatos	1,000	,929
Contenido calcio	1,000	,821
Contenido magnesio	1,000	,920
Contenido sodio	1,000	,912
Contenido potasio	1,000	,715
Contenido fosfatos	1,000	,860

Método de extracción: Análisis de Componentes principales.

Si otras variables presentan una extracción menor que 0,45 (evaluación regular) quedan en observación.
Al sacar del análisis las variables con evaluación pobre, las variables en observación pueden aumentar su poder explicativo. En caso de no hacerlo, se las elimina del análisis.



Comunalidades

	Inicial	Extracción
Conductividad eléctrica	1,000	,932
Contenido cloruros	1,000	,962
Contenido sulfatos	1,000	,905
Contenido calcio	1,000	,879
Contenido magnesio	1,000	,921
Contenido sodio	1,000	,940
Contenido potasio	1,000	,848
Contenido fosfatos	1,000	,812

Método de extracción: Análisis de Componentes principales.

Hay un ligero aumento en cloruros, calcio, magnesio, sodio y potasio.

Eliminando el contenido de bicarbonatos, los 3 primeros autovalores explican el 89,977% de la varianza, con lo que se ha obtenido un modelo más explicativo.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,986	44,291	44,291	3,986	44,291	44,291
2	1,840	20,448	64,739	1,840	20,448	64,739
3	1,549	17,207	81,946	1,549	17,207	81,946
4	,962	10,693	92,639			
5	,343	3,807	96,446			
6	,180	2,005	98,451			
7	,084	,933	99,384			
8	,041	,452	99,836			
9	,015	,164	100,000			

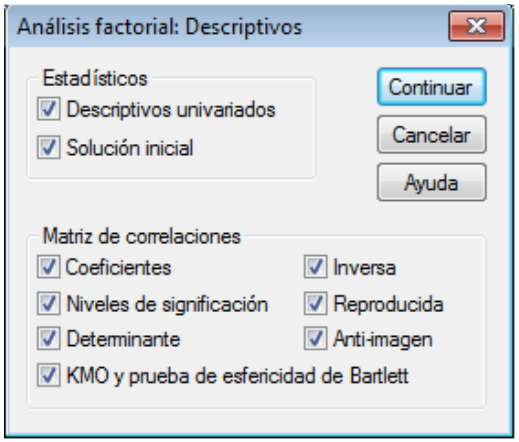
Método de extracción: Análisis de Componentes principales.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,986	49,824	49,824	3,986	49,824	49,824
2	1,668	20,848	70,672	1,668	20,848	70,672
3	1,544	19,305	89,977	1,544	19,305	89,977
4	,467	5,833	95,809			
5	,183	2,288	98,098			
6	,084	1,052	99,150			
7	,052	,644	99,794			
8	,016	,206	100,000			

Método de extracción: Análisis de Componentes principales.

Desde la matriz de correlaciones, coeficientes de correlación de Pearson entre cada par de variables, es desde donde parte el análisis.



Estadísticos descriptivos

	Media	Desviación típica	N del análisis
Conductividad eléctrica	3,8879	3,06913	17
Contenido cloruros	1160,35	1226,797	17
Contenido sulfatos	1428,53	1839,335	17
Contenido calcio	530,24	651,439	17
Contenido magnesio	165,06	230,101	17
Contenido sodio	623,53	663,392	17
Contenido potasio	13,18	11,232	17
Contenido fosfatos	5,464429	17,9034306	17

En componentes principales, la matriz de correlaciones se auto-descompone en sus autovalores y autovectores para alcanzar la solución factorial. El resto de los métodos de extracción se basan en una transformación de la matriz.

Matriz de correlaciones[a]

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación	Conductividad eléctrica	1,000	,907	,388	,380	,599	,883	,160	-,081
	Contenido cloruros	,907	1,000	,196	,255	,508	,918	,003	-,124
	Contenido sulfatos	,388	,196	1,000	,796	,866	,402	,036	-,106
	Contenido calcio	,380	,255	,796	1,000	,799	,274	,212	-,150
	Contenido magnesio	,599	,508	,866	,799	1,000	,625	,010	-,076
	Contenido sodio	,883	,918	,402	,274	,625	1,000	-,043	-,130
	Contenido potasio	,160	,003	,036	,212	,010	-,043	1,000	,630
	Contenido fosfatos	-,081	-,124	-,106	-,150	-,076	-,130	,630	1,000
Sig. (Unilateral)	Conductividad eléctrica		,000	,062	,066	,006	,000	,270	,378
	Contenido cloruros	,000		,226	,161	,019	,000	,496	,318
	Contenido sulfatos	,062	,226		,000	,000	,055	,445	,343
	Contenido calcio	,066	,161	,000		,000	,143	,207	,283
	Contenido magnesio	,006	,019	,000	,000		,004	,485	,387
	Contenido sodio	,000	,000	,055	,143	,004		,435	,310
	Contenido potasio	,270	,496	,445	,207	,485	,435		,003
	Contenido fosfatos	,378	,318	,343	,283	,387	,310	,003	

a. Determinante = 6,27E-005

Es deseable que la matriz contenga grupos de variables que correlacionen fuertemente entre sí. Una matriz de correlaciones próxima a una matriz identidad indica que el análisis factorial conducirá a una solución deficiente.

La matriz de correlaciones muestra el nivel crítico unilateral (*Sig. unilateral*) asociada a cada coeficiente de correlación (*el nivel crítico bilateral se obtiene multiplicando por 2 el unilateral*).

Un nivel crítico (*p-valor < 0,05*) rechaza la hipótesis nula, indicando que la correlación poblacional entre el correspondiente par de variables puede ser considerada significativamente distinta de 0.

Ideal Encontrar muchos niveles críticos pequeños.

Matriz de correlaciones[a]

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Sig. (Unilateral)	Conductividad eléctrica		,000	,062	,066	,006	,000	,270	,378
	Contenido cloruros	,000		,226	,161	,019	,000	,496	,318
	Contenido sulfatos	,062	,226		,000	,000	,055	,445	,343
	Contenido calcio	,066	,161	,000		,000	,143	,207	,283
	Contenido magnesio	,006	,019	,000	,000		,004	,485	,387
	Contenido sodio	,000	,000	,055	,143	,004		,435	,310
	Contenido potasio	,270	,496	,445	,207	,485	,435		,003
	Contenido fosfatos	,378	,318	,343	,283	,387	,310	,003	

a. Determinante = 6,27E-005

El valor del determinante se aproxima a 0 mostrando que las variables están linealmente relacionadas. Un buen resultado cara a la idoneidad del análisis.

Inversa de la matriz de correlaciones

	Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Conductividad eléctrica	10,953	-11,482	-4,964	3,319	-,329	2,250	-3,249	1,761
Contenido cloruros	-11,482	28,201	14,529	-10,244	-,388	-18,687	4,216	-2,557
Contenido sulfatos	-4,964	14,529	12,445	-6,145	-4,358	-9,594	1,829	-,939
Contenido calcio	3,319	-10,244	-6,145	10,670	-5,932	9,990	-4,245	3,476
Contenido magnesio	-,329	-,388	-4,358	-5,932	11,964	-3,692	3,021	-2,904
Contenido sodio	2,250	-18,687	-9,594	9,990	-3,692	19,772	-2,652	2,312
Contenido potasio	-3,249	4,216	1,829	-4,245	3,021	-2,652	3,961	-2,799
Contenido fosfatos	1,761	-2,557	-,939	3,476	-2,904	2,312	-2,799	3,094

Matriz estrechamente relacionada con la matriz *anti-imagen*. Cuando el determinante de la matriz de correlaciones vale 0, no es posible calcular la matriz inversa. En este caso tampoco es posible utilizar algunos métodos de extracción (ejes principales o máxima verosimilitud).

Los estadísticos que permiten valorar la adecuación de los datos a un modelo factorial son el KMO y la prueba de esfericidad de Bartlett.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,503
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	120,962
	gl	28
	Sig.	,000

Si KMO < 0,5 indica que el análisis factorial no es aceptable con estos datos.

La prueba de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones observada es una matriz identidad.

Se trata de comprobar que la matriz de correlaciones es significativamente distinta de la matriz identidad. Si fuera una matriz identidad no habría correlación entre las variables y no tendría sentido llevar a cabo un análisis factorial.

Se establecen las hipótesis $H_0 : |R_p| = 1 \quad H_1 : |R_p| \neq 1$

Donde n es el número de individuos de la muestra, p el número de variables utilizadas, R_p la matriz de correlaciones poblaciones y ln R el logaritmo neperiano del determinante de la matriz de correlaciones observadas: R = 0,0000627

Realiza una estimación *Chi-cuadrado* a partir de una transformación del determinante de la matriz de las correlaciones observadas.
Se calcula mediante la siguiente expresión:

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,503
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	120,962
	gl	28
	Sig.	,000

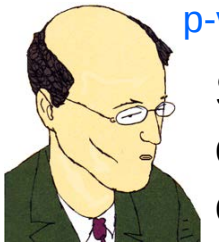
$$\chi^2 = -\left[(n-1) - \frac{1}{6}(2p+5)\right]\ln|R|$$

$$\chi^2 = -\left[(17-1) - \frac{1}{6}(2 \times 8+5)\right]\ln(0{,}0000627) = 120{,}962$$

Grados de libertad:

$$\frac{1}{2}\left(p^2 - p\right) = \frac{1}{2}\left(8^2 - 8\right) = 28$$

p-valor = 0



Si el nivel crítico (*Sig ó p-valor < 0,05*) se rechaza la hipótesis nula de esfericidad. En consecuencia, el modelo factorial es adecuado para explicar los datos.

Matriz de correlaciones[a]

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Sig. (Unilateral)	Conductividad eléctrica		,000	,062	,066	,006	,000	,270	,378
	Contenido cloruros	,000		,226	,161	,019	,000	,496	,318
	Contenido sulfatos	,062	,226		,000	,000	,055	,445	,343
	Contenido calcio	,066	,161	,000		,000	,143	,207	,283
	Contenido magnesio	,006	,019	,000	,000		,004	,485	,387
	Contenido sodio	,000	,000	,055	,143	,004		,435	,310
	Contenido potasio	,270	,496	,445	,207	,485	,435		,003
	Contenido fosfatos	,378	,318	,343	,283	,387	,310	,003	

a. Determinante = 6,27E-005

El estadístico de Kaiser-Meyer-Olkin (KMO) es un coeficiente de correlación parcial que mide la correlación existente entre dos variables una vez que se han descontado los efectos lineales de otras variables.

En un modelo factorial, los efectos lineales de otras variables se pueden interpretar como los efectos comunes. En esta línea, el coeficiente de correlación parcial entre dos variables sería equivalente al coeficiente de correlación entre los factores únicos de dos variables.

En el análisis factorial por hipótesis los coeficientes de correlación teóricos calculados entre cada par de factores son nulos.

$$KMO = \frac{\sum_{j}\sum_{h\neq j} r_{jh}^2}{\sum_{j}\sum_{h\neq j} r_{jh}^2 + \sum_{j}\sum_{h\neq j} a_{jh}^2}$$

r_{jh} son los coeficientes de correlación observados entre las variables X_j y X_h

a_{jh} son los coeficientes de correlación parcial entre las variables X_j y X_h

$$KMO = \frac{\sum_{j}\sum_{h\neq j} r_{jh}^2}{\sum_{j}\sum_{h\neq j} r_{jh}^2 + \sum_{j}\sum_{h\neq j} a_{jh}^2}$$

Cuando existe adecuación de los datos a un modelo de análisis factorial, el término del denominador que recoge los coeficientes a_{jh} es pequeño. En consecuencia, la medida KMO será próxima a 1.

Cuando KMO < 0,5 los datos son inadecuados para un análisis factorial. Mientras más cerca estén de 1 los valores de KMO mejor será la adecuación de los datos a un modelo factorial. Se considera una adecuación excelente para valores de KMO próximos a 0,9.

Existe una medida de adecuación muestral individual para cada una de las variables, basada en la medida KMO, se denomina MSA (*Measure of Sampling Adequacy*):

$$MSA_j = \frac{\sum_{h\neq j} r_{jh}^2}{\sum_{h\neq j} r_{jh}^2 + \sum_{h\neq j} a_{jh}^2}$$

Si el valor MSA_j se aproxima a 1, la variable X_j será adecuada para su tratamiento en el análisis factorial con el resto de las variables.

La matriz de correlaciones anti-imagen se utiliza como diagnóstico de la adecuación de los datos a un modelo factorial.
Un coeficiente de correlación parcial indica el grado de relación existente entre dos variables, tras eliminar el efecto de las restantes variables incluidas en el análisis.

Matrices anti-imagen

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Covarianza anti-imagen	Conductividad eléctrica	,091	-,037	-,036	,028	-,003	,010	-,075	,052
	Contenido cloruros	-,037	,035	,041	-,034	-,001	-,034	,038	-,029
	Contenido sulfatos	-,036	,041	,080	-,046	-,029	-,039	,037	-,024
	Contenido calcio	,028	-,034	-,046	,094	-,046	,047	-,100	,105
	Contenido magnesio	-,003	-,001	-,029	-,046	,084	-,016	,064	-,078
	Contenido sodio	,010	-,034	-,039	,047	-,016	,051	-,034	,038
	Contenido potasio	-,075	,038	,037	-,100	,064	-,034	,252	-,228
	Contenido fosfatos	,052	-,029	-,024	,105	-,078	,038	-,228	,323
Correlación anti-imagen	Conductividad eléctrica	,683[a]	-,653	-,425	,307	-,029	,153	-,493	,302
	Contenido cloruros	-,653	,477[a]	,776	-,591	-,021	-,791	,399	-,274
	Contenido sulfatos	-,425	,776	,513[a]	-,533	-,357	-,612	,261	-,151
	Contenido calcio	,307	-,591	-,533	,417[a]	-,525	,688	-,653	,605
	Contenido magnesio	-,029	-,021	-,357	-,525	,731[a]	-,240	,439	-,477
	Contenido sodio	,153	-,791	-,612	,688	-,240	,567[a]	-,300	,296
	Contenido potasio	-,493	,399	,261	-,653	,439	-,300	,206[a]	-,800
	Contenido fosfatos	,302	-,274	-,151	,605	-,477	,296	-,800	,240[a]

a. Medida de adecuación muestral

Un buen síntoma para aplicar el análisis factorial a los datos es que tengan una *correlación parcial pequeña* entre cualquier par de variables, propio de que las variables comparten gran cantidad de información debido a la presencia de factores comunes.

Cuando la *correlación parcial es alta* entre cualquier par de variables, comparten gran cantidad de información entre ellas, aunque no la comparten con las restantes variables, consecuentemente tampoco con los factores comunes.
Es un mal augurio para aplicar el análisis factorial a los datos.

De otra parte, las correlaciones parciales son estimaciones entre los factores únicos (*hay un factor único para cada variable del modelo*), y como deben de ser independientes entre sí, las correlaciones parciales deben ser próximas a 0.

Matrices anti-imagen

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación anti-imagen	Conductividad eléctrica	,683[a]	-,653	-,425	,307	-,029	,153	-,493	,302
	Contenido cloruros	-,653	,477[a]	,776	-,591	-,021	-,791	,399	-,274
	Contenido sulfatos	-,425	,776	,513[a]	-,533	-,357	-,612	,261	-,151
	Contenido calcio	,307	-,591	-,533	,417[a]	-,525	,688	-,653	,605
	Contenido magnesio	-,029	-,021	-,357	-,525	,731[a]	-,240	,439	-,477
	Contenido sodio	,153	-,791	-,612	,688	-,240	,567[a]	-,300	,296
	Contenido potasio	-,493	,399	,261	-,653	,439	-,300	,206[a]	-,800
	Contenido fosfatos	,302	-,274	-,151	,605	-,477	,296	-,800	,240[a]

a. Medida de adecuación muestral

En la diagonal se tiene la MSA (medida de adecuación de la muestra), con una interpretación como la del KMO, los valores han de ser cuanto más próximos a 1 mejor, valores inferiores a 0,5 son inaceptables.

Los valores de la diagonal de la matriz covarianza anti-imagen representan una estimación de la *unicidad de cada variable*, es decir, una estimación de lo que cada variable tiene de propio o de no compartido con las demás variables.

Matrices anti-imagen

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Covarianza anti-imagen	Conductividad eléctrica	,091	-,037	-,036	,028	-,003	,010	-,075	,052
	Contenido cloruros	-,037	,035	,041	-,034	-,001	-,034	,038	-,029
	Contenido sulfatos	-,036	,041	,080	-,046	-,029	-,039	,037	-,024
	Contenido calcio	,028	-,034	-,046	,094	-,046	,047	-,100	,105
	Contenido magnesio	-,003	-,001	-,029	-,046	,084	-,016	,064	-,078
	Contenido sodio	,010	-,034	-,039	,047	-,016	,051	-,034	,038
	Contenido potasio	-,075	,038	,037	-,100	,064	-,034	,252	-,228
	Contenido fosfatos	,052	-,029	-,024	,105	-,078	,038	-,228	,323

El *Visor* muestra las correlaciones reproducidas, donde se incluye la *matriz residual* donde se alojan los residuos del análisis factorial.

Cada residuo expresa la diferencia existente entre la correlación observada entre dos variables y la correlación reproducida por la estructura factorial para esas dos variables.

En una análisis factorial idóneo la mayoría de las correlaciones reproducidas se parecerán a las correlaciones observadas y los residuos serán muy pequeños.

A pie de la tabla se incluye una nota que contabiliza el número de residuos mayores que 0,05 y el porcentaje que ese número representa sobre el total de correlaciones no redundantes de la matriz.

Correlaciones reproducidas

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación reproducida	Conductividad eléctrica	,932[b]	,924	,388	,368	,633	,923	,124	-,033
	Contenido cloruros	,924	,962[b]	,236	,207	,510	,940	,001	-,118
	Contenido sulfatos	,388	,236	,905[b]	,888	,867	,359	,074	-,156
	Contenido calcio	,368	,207	,888	,879[b]	,845	,325	,153	-,074
	Contenido magnesio	,633	,510	,867	,845	,921[b]	,610	,082	-,158
	Contenido sodio	,923	,940	,359	,325	,610	,940[b]	-,020	-,166
	Contenido potasio	,124	,001	,074	,153	,082	-,020	,848[b]	,798
	Contenido fosfatos	-,033	-,118	-,156	-,074	-,158	-,166	,798	,812[b]
Residual[a]	Conductividad eléctrica		-,017	,001	,012	-,034	-,041	,036	-,048
	Contenido cloruros	-,017		-,040	,049	-,003	-,023	,002	-,005
	Contenido sulfatos	,001	-,040		-,091	-,002	,044	-,038	,050
	Contenido calcio	,012	,049	-,091		-,045	-,051	,059	-,076
	Contenido magnesio	-,034	-,003	-,002	-,045		,014	-,071	,082
	Contenido sodio	-,041	-,023	,044	-,051	,014		-,023	,036
	Contenido potasio	,036	,002	-,038	,059	-,071	-,023		-,168
	Contenido fosfatos	-,048	-,005	,050	-,076	,082	,036	-,168	

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 7 (25,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

Si hay un porcentaje elevado de diferencias superiores a 0,05 será indicativo de que el modelo factorial estimado no es adecuado para los datos.

Existen varias razones por las que la *matriz residual* podría tener un gran número de residuos altos (en valor absoluto):

- 🗁 Que se hubieran extraído un número insuficiente de factores, y en consecuencia la estructura factorial no fuera capaz de reproducir adecuadamente la matriz de correlaciones observada.
- 🗁 Las correlaciones observadas podrían estar mal estimadas, por la presencia de sesgos en la medida de las variables, o porque el coeficiente de correlación de Pearson no fuera apropiado por la escala utilizada para medir las variables.

Desde la matriz de componentes con las correlaciones de las variables en cada componente, se calculan las correlaciones reproducidas y las correlaciones residuales.

Matriz de componentes[a]

	Componente		
	1	2	3
Conductividad eléctrica	,867	-,091	,414
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido magnesio	,894	,161	-,309
Contenido sodio	,858	-,246	,379
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

Matriz de componentes con las correlaciones de las variables en cada componente.

Así, por ejemplo, la correlación reproducida entre el contenido de magnesio y conductividad eléctrica se obtiene sumando los productos de las correlaciones entre las dos variables.

$$\text{Correlación reproducida} = (0{,}894 \times 0{,}867) - (0{,}161 \times 0{,}091) - (0{,}309 \times 0{,}414) = 0{,}633$$

Correlaciones reproducidas

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación reproducida	Conductividad eléctrica	,932[b]	,924	,388	,368	,633	,923	,124	-,033
	Contenido cloruros	,924	,962[b]	,236	,207	,510	,940	,001	-,118
	Contenido sulfatos	,388	,236	,905[b]	,888	,867	,359	,074	-,156
	Contenido calcio	,368	,207	,888	,879[b]	,845	,325	,153	-,074
	Contenido magnesio	,633	,510	,867	,845	,921[b]	,610	,082	-,158
	Contenido sodio	,923	,940	,359	,325	,610	,940[b]	-,020	-,166
	Contenido potasio	,124	,001	,074	,153	,082	-,020	,848[b]	,798
	Contenido fosfatos	-,033	-,118	-,156	-,074	-,158	-,166	,798	,812[b]

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 7 (25,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

La matriz de correlaciones residual es la diferencia entre las correlaciones observadas y las reproducidas.

Matriz de correlaciones[a]

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación	Conductividad eléctrica	1,000	,907	,388	,380	,599	,883	,160	-,081
	Contenido cloruros	,907	1,000	,196	,255	,508	,918	,003	-,124
	Contenido sulfatos	,388	,196	1,000	,796	,866	,402	,036	-,106
	Contenido calcio	,380	,255	,796	1,000	,799	,274	,212	-,150
	Contenido magnesio	,599	,508	,866	,799	1,000	,625	,010	-,076
	Contenido sodio	,883	,918	,402	,274	,625	1,000	-,043	-,130
	Contenido potasio	,160	,003	,036	,212	,010	-,043	1,000	,630
	Contenido fosfatos	-,081	-,124	-,106	-,150	-,076	-,130	,630	1,000

a. Determinante = 6,27E-005

Correlaciones reproducidas

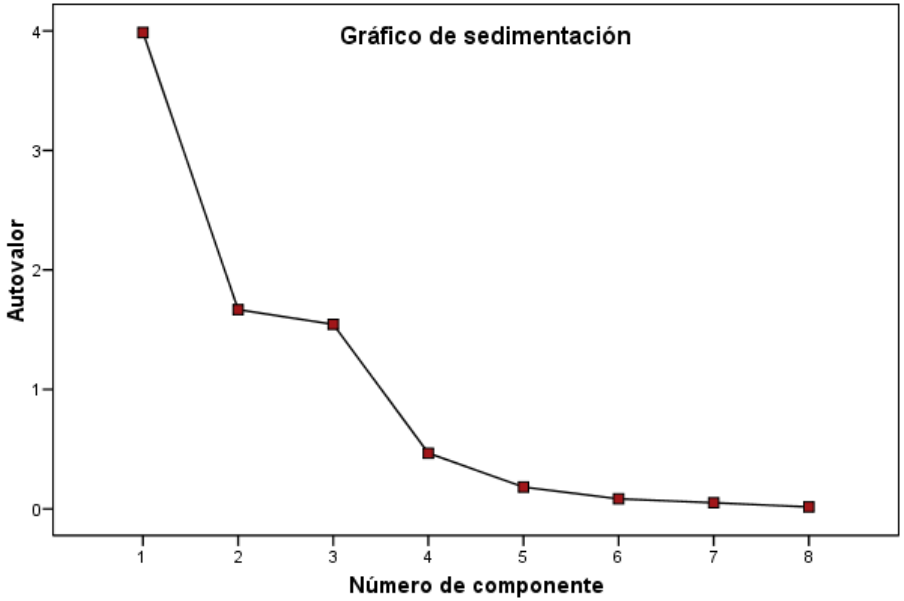
		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Correlación reproducida	Conductividad eléctrica	,932[b]	,924	,388	,368	,633	,923	,124	-,033
	Contenido cloruros	,924	,962[b]	,236	,207	,510	,940	,001	-,118
	Contenido sulfatos	,388	,236	,905[b]	,888	,867	,359	,074	-,156
	Contenido calcio	,368	,207	,888	,879[b]	,845	,325	,153	-,074
	Contenido magnesio	,633	,510	,867	,845	,921[b]	,610	,082	-,158
	Contenido sodio	,923	,940	,359	,325	,610	,940[b]	-,020	-,166
	Contenido potasio	,124	,001	,074	,153	,082	-,020	,848[b]	,798
	Contenido fosfatos	-,033	-,118	-,156	-,074	-,158	-,166	,798	,812[b]

Correlaciones reproducidas

		Conductividad eléctrica	Contenido cloruros	Contenido sulfatos	Contenido calcio	Contenido magnesio	Contenido sodio	Contenido potasio	Contenido fosfatos
Residual[a]	Conductividad eléctrica		-,017	,001	,012	-,034	-,041	,036	-,048
	Contenido cloruros	-,017		-,040	,049	-,003	-,023	,002	-,005
	Contenido sulfatos	,001	-,040		-,091	-,002	,044	-,038	,050
	Contenido calcio	,012	,049	-,091		-,045	-,051	,059	-,076
	Contenido magnesio	-,034	-,003	-,002	-,045		,014	-,071	,082
	Contenido sodio	-,041	-,023	,044	-,051	,014		-,023	,036
	Contenido potasio	,036	,002	-,038	,059	-,071	-,023		-,168
	Contenido fosfatos	-,048	-,005	,050	-,076	,082	,036	-,168	

Es importante señalar que el gráfico de sedimentaciones no varía con el número de factores seleccionado.

El gráfico siempre muestra todos los posibles autovalores de la matriz de correlación original y no los autovalores de la matriz analizada.



Con la matriz de las componentes se calculan las *comunalidades*, sumando los coeficientes al cuadrado de cada variable en los factores extraídos.

Matriz de componentes[a]

	Componente		
	1	2	3
Conductividad eléctrica	,867	-,091	,414
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido magnesio	,894	,161	-,309
Contenido sodio	,858	-,246	,379
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

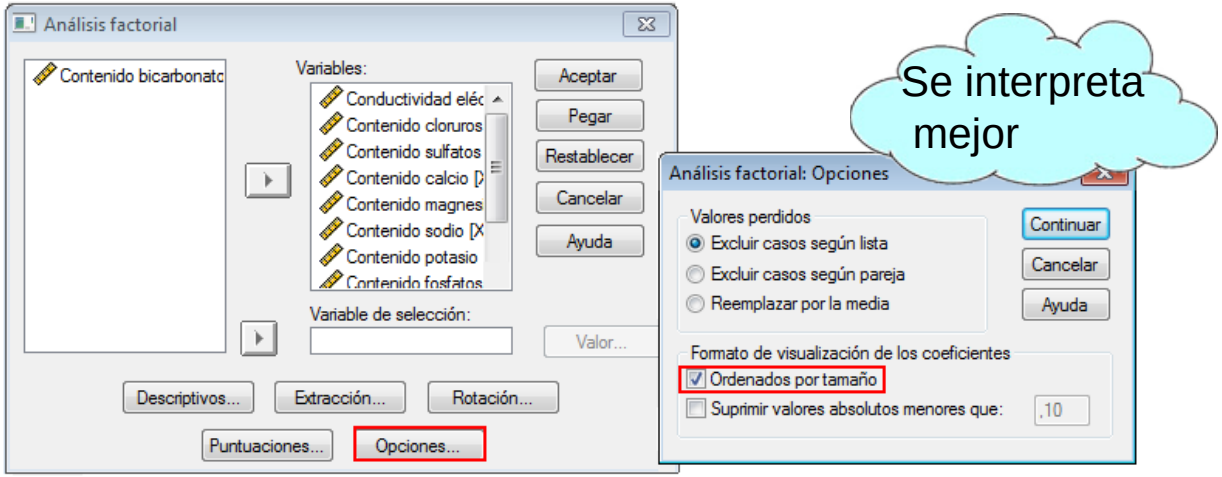
Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

Comunalidades

	Inicial	Extracción
Conductividad eléctrica	1,000	,932
Contenido cloruros	1,000	,962
Contenido sulfatos	1,000	,905
Contenido calcio	1,000	,879
Contenido magnesio	1,000	,921
Contenido sodio	1,000	,940
Contenido potasio	1,000	,848
Contenido fosfatos	1,000	,812

Método de extracción: Análisis de Componentes principales.



Se interpreta mejor

Matriz de componentes[a]

	Componente		
	1	2	3
Contenido magnesio	,894	,161	-,309
Conductividad eléctrica	,867	-,091	,414
Contenido sodio	,858	-,246	,379
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

$$\text{Comunalidad} = F_1^2 + F_2^2 + F_3^2$$

$\mapsto$

Comunalidades

	Inicial	Extracción
Conductividad eléctrica	1,000	,932
Contenido cloruros	1,000	,962
Contenido sulfatos	1,000	,905
Contenido calcio	1,000	,879
Contenido magnesio	1,000	,921
Contenido sodio	1,000	,940
Contenido potasio	1,000	,848
Contenido fosfatos	1,000	,812

Método de extracción: Análisis de Componentes principales.

$$\text{Comunalidad magnesio} = 0{,}894^2 + 0{,}161^2 + (-0{,}309)^2 = 0{,}921$$

Peso o *saturación factorial* es la cifra resultante de unir la fila de cada variable con la columna de cada factor

La situación idónea es que todas las variables tengan pesos o saturaciones altos en un factor y bajos en los demás factores.

Matriz de componentes[a]

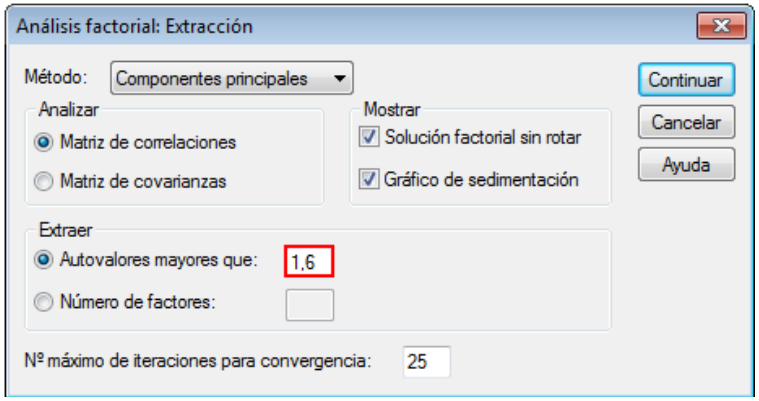
	Componente		
	1	2	3
Contenido magnesio	,894	,161	-,309
Conductividad eléctrica	,867	-,091	,414
Contenido sodio	,858	-,246	,379
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

La variable X_6 = "contenido magnesio" tiene una relación 0,894 con el primer factor, de 0,161 con el segundo, y de –0,309 con el tercero; de modo que X_6 forma parte del primer factor.

La variable X_8 = "contenido potasio" tiene una relación 0,062 con el primer factor, de 0,847 con el segundo, y de 0,356 con el tercero; de modo que X_8 forma parte del segundo factor.



Adviértase que en la extracción se había fijado el criterio de extraer autovalores mayores que 1.
Si se hubiera decidido que fueran mayor que 1,6 se tendría:

Matriz de componentes[a]

	Componente	
	1	2
Contenido magnesio	,894	,161
Conductividad eléctrica	,867	-,091
Contenido sodio	,858	-,246
Contenido cloruros	,794	-,270
Contenido sulfatos	,734	,262
Contenido calcio	,705	,345
Contenido potasio	,062	,847
Contenido fosfatos	-,165	,771

Observaciones

Una variable puede tener menor comunalidad por su permanencia simultánea a dos factores, en contra de lo ideal que estuviera saturada en un único factor.
Es una *variable compleja* que no sirve para identificar la naturaleza de los factores.
Lo mejor es eliminarla del modelo, lo que produciría un aumento en el porcentaje de varianza explicada.

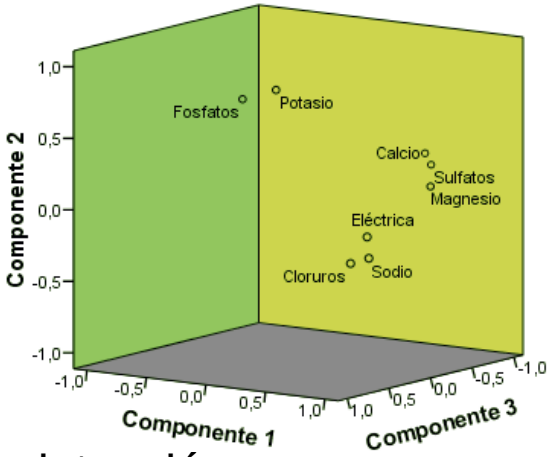
Matriz de componentes[a]

	Componente		
	1	2	3
Contenido magnesio	,894	,161	-,309
Conductividad eléctrica	,867	-,091	,414
Contenido sodio	,858	-,246	,379
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

SPSS muestra el gráfico de saturación en donde se recogen los valores de la matriz de componentes, al tratarse de una representación tridimensional no suele ser de mucha utilidad.



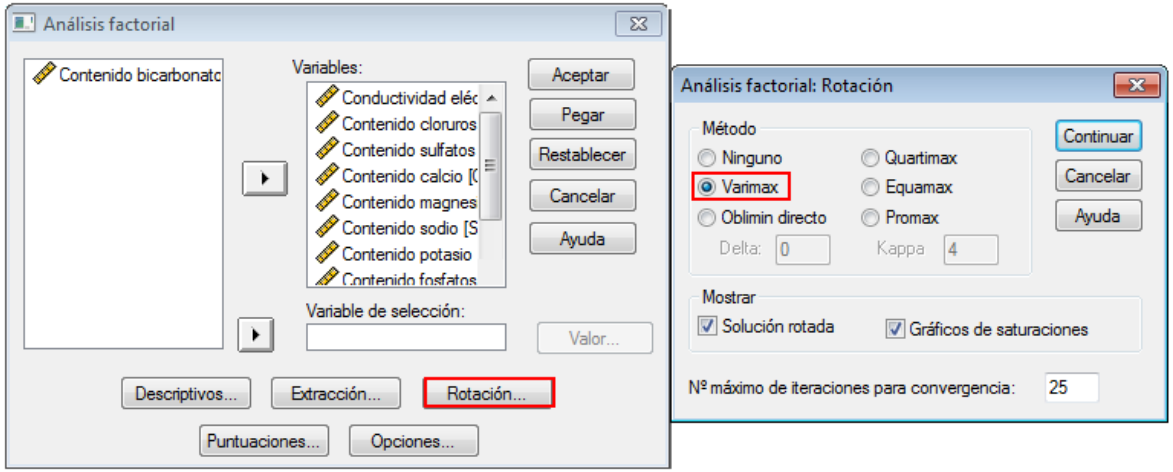
Realizado el ajuste de datos, se procede a interpretar la matriz factorial (componentes) que indica la relación entre los factores y variables.

En muchas ocasiones no es una tarea fácil, teniendo que recurrir a *rotar* o *girar* los ejes de coordenadas que representan los factores, con la intención de que los ejes se aproximen a las variables donde están saturados.

La configuración de una estructura factorial no es única, cualquier solución factorial puede transformarse en otra.

Con nuevas referencias los pesos factoriales pueden ser más comprensibles y se pueda interpretar mejor el modelo.

Con la rotación de ejes sólo cambia la varianza explicada por cada factor. Las comunalidades y el porcentaje de varianza explicado queda inalterable.



Hay diversos métodos para realizar rotaciones factoriales, divididos en dos grandes grupos: métodos ortogonales y oblicuos.

- Método ortogonal: Gira los ejes ortogonalmente, en el mismo ángulo, motivo por el que se utiliza cuando no existe relación entre los factores. Los más importantes:

  - Varimax: Minimiza el número de variables que tienen saturaciones altas en cada factor. El objetivo es aumentar las saturaciones más altas en un factor, mientras que se disminuyen las más bajas para que el factor se pueda interpretar mejor. Es el método más utilizado en la investigación de mercado y social.

🗎 Quartimax: Minimiza el número de factores necesarios para explicar cada variable, concentrando la mayor parte de la varianza de cada variable en un factor y dejando próximas a 0 el resto de saturaciones de cada variable.

🗎 Equamax: Combinación del método *Varimax* que simplifica los factores y el método *Quartimax* que simplifica las variables.

- Método oblicuo: Considera que dos factores pueden explicar una misma realidad, que existe correlación entre los factores, y en consecuencia cada eje podría girar en un ángulo diferente. Se utiliza poco en la investigación de mercado y social.

🗎 Oblimin directo: Cuando *Delta* es igual a 0 (valor por defecto) las soluciones son las más oblicuas. A medida que *Delta* se va haciendo más negativo los factores son menos oblicuos. Para anular el valor por defecto de *Delta* se puede introducir un número menor o igual que 0,8.

🗎 Promax: Se calcula más rápido que una rotación *Oblimin directo*, por lo que es útil para un gran conjunto de datos. El parámetro que controla el cálculo es *Kappa*, el valor por defecto es 4, valor válido para la mayoría de los análisis.

Comunalidades y porcentaje de varianza explicado permanecen invariantes.

La Comunalidad es el porcentaje de varianza de cada variable que explica el análisis factorial.

Con la rotación de ejes cambia la matriz de las componentes (relación entre los factores y las variables), es decir, la influencia de cada variable en el factor.

Matriz de componentes[a]

	Componente		
	1	2	3
Contenido magnesio	,894	,161	-,309
Conductividad eléctrica	,867	-,091	,414
Contenido sodio	,858	-,246	,379
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,513
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

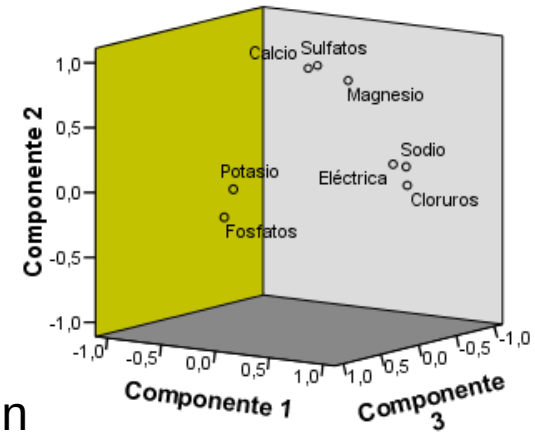
Matriz de componentes rotados[a]

	Componente		
	1	2	3
Contenido cloruros	,975	,099	-,043
Contenido sodio	,937	,234	-,082
Conductividad eléctrica	,923	,273	,071
Contenido sulfatos	,145	,939	-,039
Contenido calcio	,120	,929	,049
Contenido magnesio	,435	,855	-,030
Contenido potasio	,030	,112	,914
Contenido fosfatos	-,070	-,118	,890

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Gráfico de saturación de componentes con *rotación varimax*



Con la *rotación varimax* las variables tienen saturaciones más altas en un factor y más bajas en los demás factores.

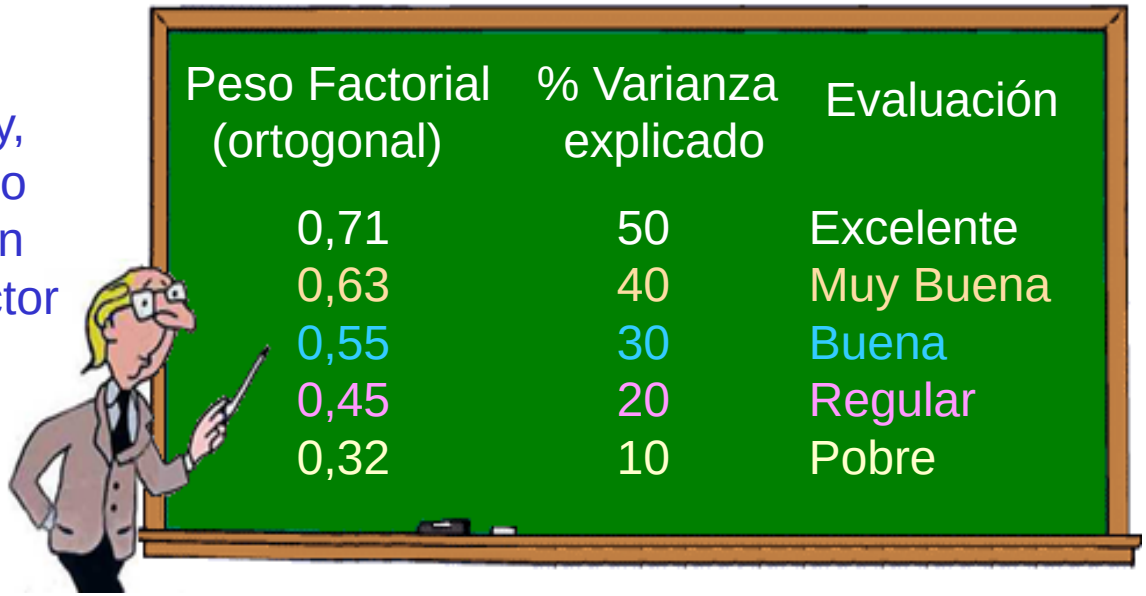
La variable X_3= “contenido cloruros” tiene una relación 0,975 con el primer factor, de 0,099 con el segundo, y de –0,043 con el tercero; de modo que X_3 forma parte del primer factor.

La variable X_4 = “contenido sulfatos” tiene una relación 0,145 con el primer factor, de 0,939 con el segundo, y de –0,039 con el tercero; de modo que X_4 forma parte del segundo factor.

La variable X_8 = “contenido potasio” tiene una relación 0,030 con el primer factor, de 0,112 con el segundo, y de 0,914 con el tercero; de modo que X_8 forma parte del tercer factor.

Un factor se interpreta fácilmente cuando varias variables correlacionan altamente con éste, y además no están relacionadas con otros factores.

Tabla de correlaciones variable-factor de Comrey, indicando hasta qué punto es adecuada la saturación de una variable en un factor

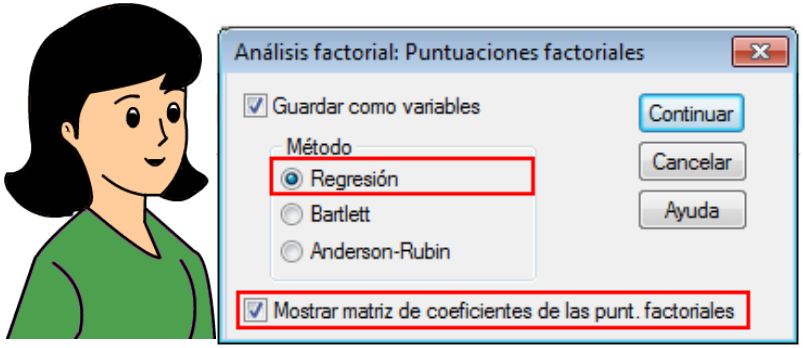


Peso Factorial (ortogonal)	% Varianza explicado	Evaluación
0,71	50	Excelente
0,63	40	Muy Buena
0,55	30	Buena
0,45	20	Regular
0,32	10	Pobre

Se consideran únicamente los coeficientes mayores de 0,30, pues un coeficiente de 0,32 indica que sólo un 10% de la variable es explicada por el factor, mientras que el resto son factores de error, específicos y comunes.

Encontrada la solución factorial final, es importante obtener una estimación de las puntuaciones de los sujetos en cada uno de los factores (componentes) resultantes de la extracción, con el fin de valorar la situación relativa de cada sujeto en las *nuevas dimensiones ocultas* capaces de resumir la información contenida en las variables originales.

- Guardar como variables: En el Editor de datos se guardan automáticamente las puntuaciones factoriales estimadas para cada sujeto en cada uno de los factores (componentes) obtenidos en la solución factorial.
El SPSS crea en el *archivo de datos* tantas variables nuevas como factores (componentes) contenga la solución final, tipo (FAC1_1, FAC2_1, …)



En el apartado *Método* hay varias técnicas de estimación de las puntuaciones factoriales, siendo *Regresión* la más generalizada.

Cuando se selecciona *componentes principales* como método de extracción, las opciones del recuadro no tienen efecto alguno, ya que en este modelo factorial las puntuaciones factoriales no son estimadas, se calculan directamente a partir de las variables originales.

- Regresión: Método de estimación de puntuaciones factoriales, donde las estimaciones resultantes tienen media 0 y como varianza el cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Con este método las puntuaciones factoriales estimadas pueden estar correlacionadas incluso cuando los factores son ortogonales.

- Bartlett: Método de estimación de puntuaciones factoriales, donde las estimaciones resultantes tienen media 0. Minimiza la suma de cuadrados de los factores únicos, es decir, minimiza la unicidad correspondiente a cada una de las variables incluidas en el análisis.

- Anderson-Rubin: Es una modificación del método de Bartlett que asegura la ortogonalidad de las puntuaciones factoriales estimadas. Las estimaciones realizadas tienen media 0, desviación típica 1, siendo independientes entre sí.

- Mostrar matriz de coeficientes de las puntuaciones factoriales: Tabla con los pesos o ponderaciones necesarios para calcular las puntuaciones factoriales a partir de las variables originales.

Matriz de coeficientes para el cálculo de puntuaciones factoriales.

Al utilizar el método de extracción de componentes principales, las dimensiones obtenidas reciben el nombre de componentes en lugar de factores.

Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente		
	1	2	3
Conductividad eléctrica	,340	-,050	,063
Contenido cloruros	,393	-,141	-,004
Contenido sulfatos	-,120	,410	-,025
Contenido calcio	-,127	,409	,028
Contenido magnesio	,019	,315	-,013
Contenido sodio	,350	-,071	-,029
Contenido potasio	,013	,042	,556
Contenido fosfatos	,014	-,046	,541

Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.
Puntuaciones de componentes.

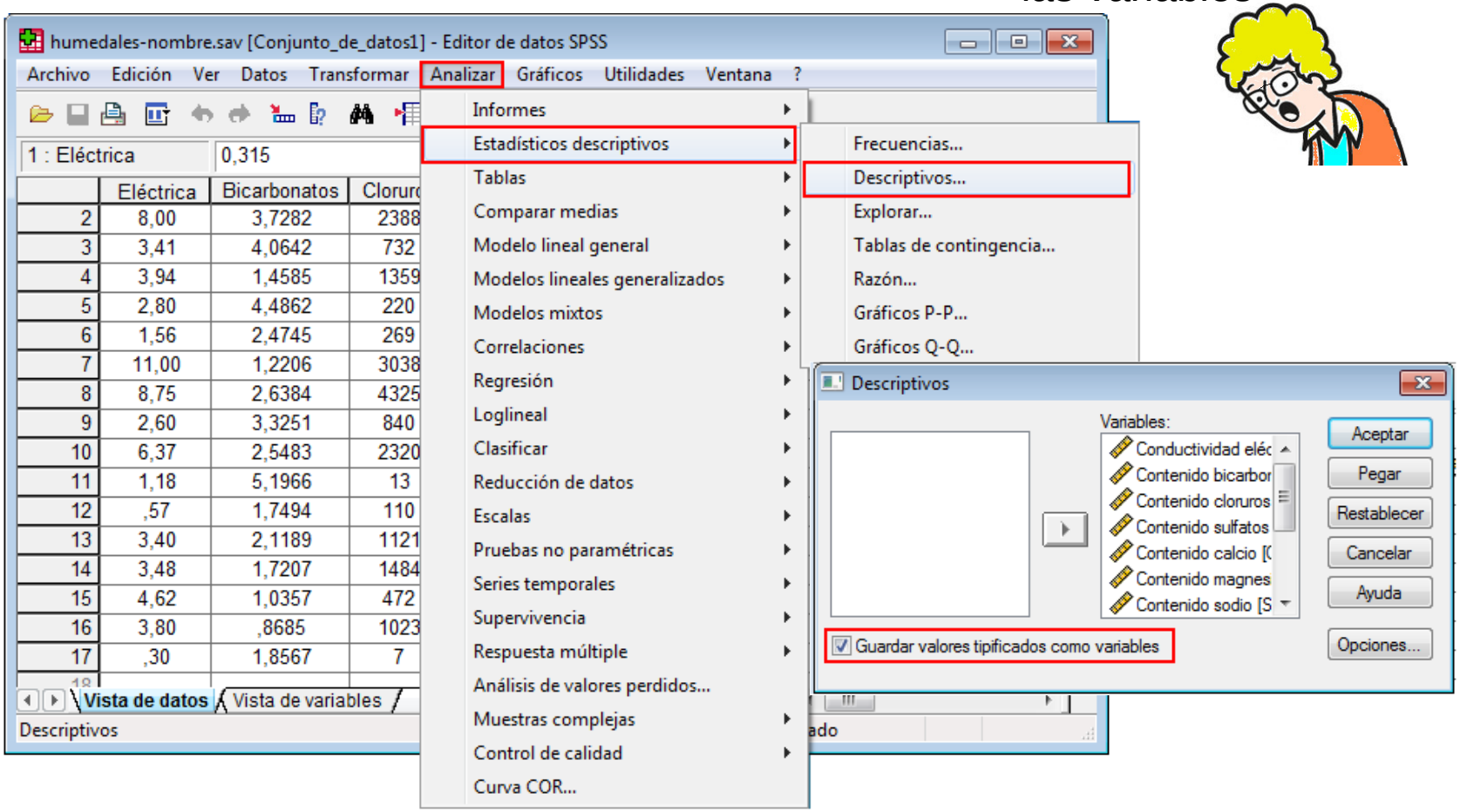
Las ecuaciones lineales en las que se basa el cálculo de puntuaciones factoriales, se elaboran combinando cada variable tipificada con su correspondiente coeficiente:

$$Y_1 = 0{,}340 \times ZEléctrica + 0{,}393 \times ZCloruros - 0{,}120 \times ZSulfatos - 0{,}127 \times ZCalcio + \\ + 0{,}019 \times ZMagnesio + 0{,}350 \times ZSodio + 0{,}013 \times ZPotasio + 0{,}014 \times ZFosfatos$$

$$Y_2 = -0{,}050 \times ZEléctrica - 0{,}141 \times ZCloruros + 0{,}410 \times ZSulfatos + 0{,}409 \times ZCalcio + \\ + 0{,}315 \times ZMagnesio - 0{,}071 \times ZSodio + 0{,}042 \times ZPotasio - 0{,}046 \times ZFosfatos$$

$$Y_3 = 0{,}063 \times z\text{eléctrica} - 0{,}004 \times z\text{cloruros} - 0{,}025 \times z\text{sulfatos} + 0{,}028 \times z\text{calcio} -$$
$$- 0{,}013 \times z\text{magnesio} - 0{,}029 \times z\text{sodio} + 0{,}556 \times z\text{potasio} + 0{,}541 \times z\text{fosfatos}$$

Se tipifican
las variables



En el Editor de datos aparecen las variables tipificadas: ZEléctrica, ZCloruros, ZSulfatos, ZCalcio, ZMagnesio, ZSodio, ZPotasio, ZFosfatos

*humedales-nombre.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

Visible: 20 de 20 variable

	ZEléctrica	ZCloruros	ZSulfatos	ZCalcio	ZMagnesio	ZSodio	ZPotasio	ZFosfatos
1	-1,16416	-,94176	-,72990	-,72952	-,69995	-,93388	-,99507	-,20422
2	1,33981	1,00069	3,37593	2,44499	3,50690	1,70860	-,72797	-,28719
3	-,15573	-,34916	-,29768	-,47930	-,18713	-,36860	2,47720	3,86091
4	,01696	,16192	-,35694	-,42864	-,30012	,13035	,07332	-,29981
5	-,35448	-,76651	,587				-,54991	-,29981
6	-,75850	-,72657	-,507				-,37184	-,28719
7	2,31729	1,53053	-,274				-,19378	-,30161
8	1,58418	2,57960	-,528				-,19378	-,23849
9	-,41964	-,26113	,457				-,54991	-,27636
10	,80872	,94526	-,211				,87461	-,27817
11	-,88232	-,93524	-,505				,42945	,07352
12	-1,08107	-,85617	-,753				-,63894	-,27275
13	-,15898	-,03208	-,305				-,81700	-,26374
14	-,13292	,26381	-,475				-,54991	-,29801
15	,23852	-,56110	,83480	,34042	-,02199	,04292	1,85397	-,29801
16	-,02865	-,11196	,45966	2,17329	,84720	-,29173	,87461	-,27456
17	-1,16904	-,94013	-,76850	-,75408	-,70429	-,93388	-,99507	-,05453
18								

	Componente		
	1	2	3
Conductividad eléctrica	,340	-,050	,063
Contenido cloruros	,393	-,141	-,004
Contenido sulfatos	-,120	,410	-,025
Contenido calcio	-,127	,409	,028
Contenido magnesio	,019	,315	-,013
Contenido sodio	,350	-,071	-,029
Contenido potasio	,013	,042	,556
Contenido fosfatos	,014	-,046	,541

Vista de datos | Vista de variables

SPSS El procesador está preparado

$$Y_1 = -0,340 \times 1,16416 - 0,393 \times 0,94176 + 0,120 \times 0,72990 + 0,127 \times 0,72990 -$$
$$- 0,019 \times 0,6995 - 0,350 \times 0,93388 - 0,013 \times 0,99507 - 0,014 \times 0,20422 = -0,94151$$

$$Y_2 = 0,340 \times 1,33981 + 0,393 \times 1,00069 - 0,120 \times 3,37593 - 0,127 \times 2,44499 +$$
$$+ 0,019 \times 3,50690 + 0,350 \times 1,70860 - 0,013 \times 0,72797 - 0,014 \times 0,28719 = 0,78748$$

Visor SPSS

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ventana ?

- Informes
- Estadísticos descriptivos
- Tablas
- Comparar medias
- Modelo lineal general
- Modelos lineales gene
- Modelos mixtos
- Correlaciones
- Regresión
- Loglineal
- Clasificar
- Reducción de datos
- Escalas
- Pruebas no paramétric
- Series temporales
- Supervivencia
- Respuesta múltiple
- Análisis de valores per
- Muestras complejas
- Control de calidad
- Curva COR...

- Cubos OLAP...
- Resúmenes de casos...
- Informe de estadísticos en filas...
- Informe de estadísticos en columnas...

Matriz de componentes[a]

	Componente		
	1	2	3
Conductividad eléctrica	,867	-,091	,414
Contenido cloruros	,794	-,270	,508
Contenido sulfatos	,734	,262	-,544
Contenido calcio	,705	,345	-,51
Contenido magnesio	,894	,161	-,30
Contenido sodio	,858	-,246	,379
Contenido potasio	,062	,847	,356
Contenido fosfatos	-,165	,771	,435

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

Matriz de componentes rotados[a]

	Componente		
	1	2	3
Conductividad eléctrica	,923	,273	,071
Contenido cloruros	,975	,099	-,04
Contenido sulfatos	,145	,939	-,039
Contenido calcio	,120	,929	,049
Contenido magnesio	,435	,855	-,030
Contenido sodio	,937	,234	-,082
Contenido potasio	,030	,112	,914
Contenido fosfatos	-,070	-,118	,890

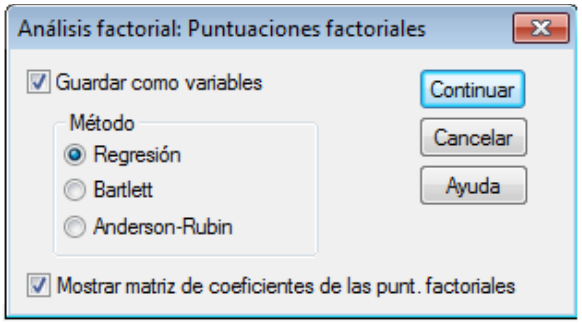
Método de extracción: Análisis de componentes principales.
Método de rotación: Normalización Varimax con Kaiser.

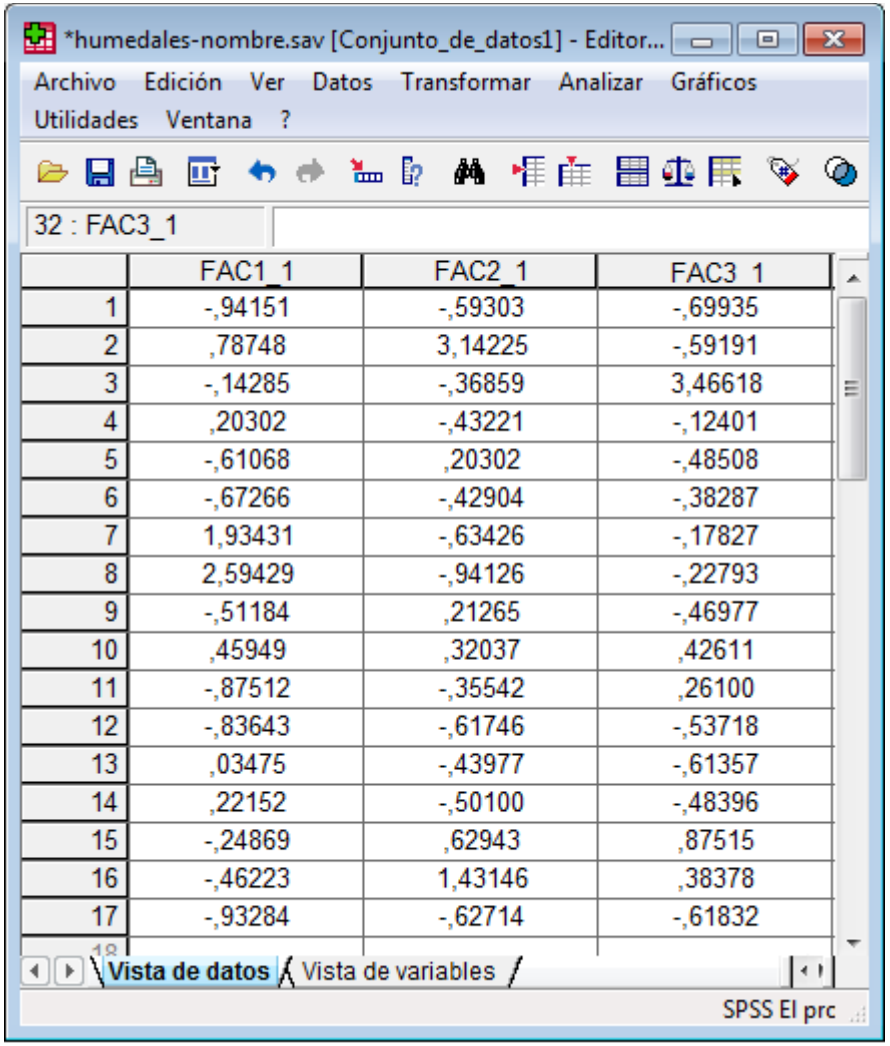
a. La rotación ha convergido en 5 iteraciones.

Resúmenes de casos

		REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1	REGR factor score 3 for analysis 1
1		-,94151	-,59303	-,69935
2		,78748	3,14225	-,59191
3		-,14285	-,36859	3,46618
4		,20302	-,43221	-,12401
5		-,61068	,20302	-,48508
6		-,67266	-,42904	-,38287
7		1,93431	-,63426	-,17827
8		2,59429	-,94126	-,22793
9		-,51184	,21265	-,46977
10		,45949	,32037	,42611
11		-,87512	-,35542	,26100
12		-,83643	-,61746	-,53718
13		,03475	-,43977	-,61357
14		,22152	-,50100	-,48396
15		-,24869	,62943	,87515
16		-,46223	1,43146	,38378
17		-,93284	-,62714	-,61832
Total	Media	,0000000	,0000000	,0000000
	Varianza	1,000	1,000	1,000

Para calcular las puntuaciones factoriales con SPSS basta con marcar la opción *Guardar como variables*. En el *Editor de datos* aparecen las puntuaciones factoriales de cada uno de los sujetos en los factores extraídos: FAC1_1, FAC2_1 y FAC3_1.





	FAC1_1	FAC2_1	FAC3_1
1	-,94151	-,59303	-,69935
2	,78748	3,14225	-,59191
3	-,14285	-,36859	3,46618
4	,20302	-,43221	-,12401
5	-,61068	,20302	-,48508
6	-,67266	-,42904	-,38287
7	1,93431	-,63426	-,17827
8	2,59429	-,94126	-,22793
9	-,51184	,21265	-,46977
10	,45949	,32037	,42611
11	-,87512	-,35542	,26100
12	-,83643	-,61746	-,53718
13	,03475	-,43977	-,61357
14	,22152	-,50100	-,48396
15	-,24869	,62943	,87515
16	-,46223	1,43146	,38378
17	-,93284	-,62714	-,61832

Las puntuaciones factoriales se encuentran en formato *diferencial*, por lo que una puntuación de 0 se corresponde con una puntuación factorial igual a la media, las puntuaciones positivas son puntuaciones mayores que la media y las puntuaciones negativas son puntuaciones menores que la media.

El análisis factorial puede verse como un análisis de regresión múltiple en donde una serie de variables independientes tratan de explicar una variable dependiente o factor (componente).

La diferencia se encuentra en que en la *regresión múltiple* el investigador decide la variable dependiente, mientras que en el *análisis factorial* la variable dependiente es una variable latente que es creada por la combinación lineal de éstas.

Con las puntuaciones factoriales de cada unos de los sujetos en los factores (componentes) extraídos, rotuladas con los nombres FAC1_1, FAC2_1, FAC3_1, se puede trabajar igual que con el resto de variables de la investigación, sabiendo que se trata de variables tipificadas.

De este modo, se puede realizar un *análisis cluster* tratando de clasificar los sujetos en grupos según sea su comportamiento, tratando de maximizar la varianza entre los grupos y minimizar la varianza dentro del grupo.