



**KIWITEC.**  
HIGH QUALITY TECH COURSES

Curso:  
**SPSS STATISTICS**

Módulo I:  
**ANÁLISIS  
FACTORIAL**

[www.kiwitec.es](http://www.kiwitec.es) - [secretaria@kiwitec.es](mailto:secretaria@kiwitec.es)





# I. Introducción

El análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de ellas. Esos grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

Cuando recogemos un gran número de variables de forma simultánea, como por ejemplo en un cuestionario de satisfacción laboral, podemos estar interesados en averiguar si las preguntas del cuestionario se agrupan de alguna forma característica. Aplicando un análisis factorial a las respuestas de los sujetos podemos encontrar grupos de variables con significado común y conseguir de esta manera reducir el número de dimensiones necesarias para explicar las respuestas de los sujetos.

El análisis factorial es, por tanto, una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos.

A diferencia de lo que ocurre en otras técnicas como el análisis de varianza o el de regresión, en el análisis factorial todas las variables del análisis cumplen el mismo papel: todas ellas son independientes en el sentido de que no existe a priori una dependencia conceptual de unas variables sobre otras.

# II. Análisis factorial

El análisis factorial consta de **cuatro fases** características:

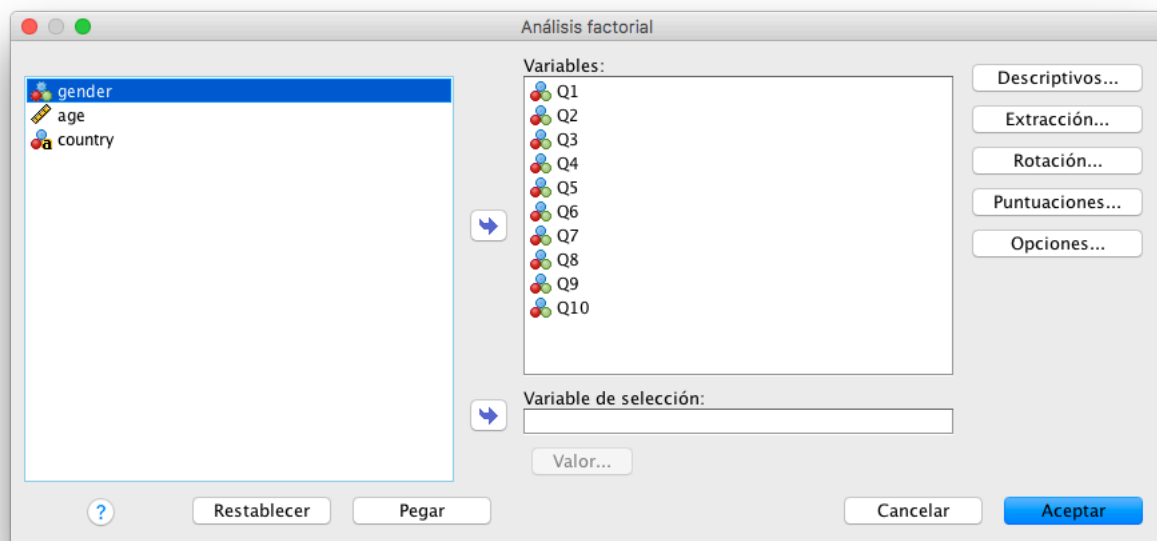
- El **cálculo de una matriz** capaz de expresar la variabilidad conjunta de todas las variables.
- La **extracción** del número óptimo de **factores**.
- La **rotación de la solución** para facilitar su interpretación.
- La **estimación de las puntuaciones de los sujetos** en las nuevas dimensiones.

Para ejecutar correctamente un análisis factorial será necesario tomar algunas decisiones en cada una de estas fases. La estructura del procedimiento Análisis factorial del SPSS se ajusta a las cuatro fases mencionadas.

Para llevar a cabo un análisis factorial hay que seleccionar la opción **Reducción de datos > Análisis factorial...** del menú Analizar:



La lista de variables contiene un listado de todas las variables del archivo, incluidas las variables de cadena (aunque éstas sólo pueden utilizarse como variables de selección). Debe seleccionarse el conjunto de variables que se desea analizar y trasladarlas a la lista **Variables**.



El cuadro **Variable de selección** permite seleccionar una de las variables del archivo de datos como variable de filtro: para definir una sub-muestra de sujetos que cumplan una determinada condición. Esta opción es especialmente útil cuando se ha reservado un porcentaje de los sujetos de la muestra para llevar a cabo una validación cruzada del modelo final. Para utilizar una variable de selección debe trasladarse la variable al cuadro **Variable de selección** y pulsar en el botón **Valor...** para acceder a un subcuadro de diálogo donde se introducirá el valor de la variable de selección que identifica a los casos que se desea incluir en el análisis. (Para aprovechar al máximo las posibilidades de esta opción, puede utilizarse el proceso *Ejecutar casos no seleccionados*, el cual actúa sobre la tabla de notas del análisis conmutando los casos seleccionados por los no seleccionados y volviendo a ejecutar el procedimiento **Análisis factorial** con las mismas especificaciones establecidas en el análisis



precedente).

Al ejecutar el análisis con las opciones por defecto se muestra una tabla con las comunales asignadas inicialmente a las variables (inicial) y las comunales reproducidas por la solución factorial (extracción).

#### Comunalidades

	Inicial	Extracción
Q1	1,000	,713
Q2	1,000	,696
Q3	1,000	,633
Q4	1,000	,595
Q5	1,000	,586
Q6	1,000	,679
Q7	1,000	,638
Q8	1,000	,545
Q9	1,000	,696
Q10	1,000	,740

Método de extracción:  
análisis de componentes  
principales.

**La comunalidad de una variable es la proporción de su varianza que puede ser explicada por el modelo factorial obtenido.** Estudiando las comunales de la extracción podemos valorar cuáles de las variables son peor explicadas por el modelo.

El método de extracción utilizado por defecto es el método de componentes principales. Dicho método de extracción, asume que es posible explicar el 100% de la varianza observada y, por ello, todas las comunales iniciales son iguales a la unidad.

A partir de esta tabla puede comenzar a determinarse si el número de factores obtenidos es suficiente para explicar todas y cada una de las variables incluidas en el análisis, así como si todas las variables pueden formar parte del análisis o alguna debe quedarse fuera.

Posteriormente se muestra la tabla de porcentajes de *varianza explicada*, con un listado de los autovalores de la matriz de varianzas-covarianzas y del porcentaje de varianza que representa cada uno de ellos.

Los **autovalores** expresan la cantidad de la varianza total que está explicada por cada factor; y los porcentajes de varianza explicada asociados a cada factor se obtienen dividiendo su correspondiente autovalor por la suma de los autovalores (que coincide con el número de variables). Por defecto, se extraen tantos factores como autovalores mayores que 1 tiene la matriz analizada.

#### Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	5,425	54,254	54,254	5,425	54,254	54,254
2	1,095	10,951	65,205	1,095	10,951	65,205
3	,627	6,267	71,472			
4	,576	5,765	77,237			
5	,523	5,234	82,471			
6	,498	4,981	87,451			
7	,376	3,758	91,210			
8	,337	3,366	94,576			
9	,279	2,789	97,365			
10	,264	2,635	100,000			

Método de extracción: análisis de componentes principales.



La tabla muestra también, para cada factor con autovalor mayor que 1, la suma de las saturaciones al cuadrado. Las sumas de cuadrados de la columna Total (que coinciden con los autovalores cuando se utiliza el método *componentes principales*, pero no cuando se utilizan otros métodos de extracción), pueden ayudar a determinar el número idóneo de factores, en función del porcentaje de varianza que se quiera explicar.

### Matriz de componente

	Componente	
	1	2
Q1	,760	,367
Q2	,681	,481
Q3	-,777	,172
Q4	,637	,435
Q5	-,751	,150
Q6	,815	,123
Q7	,791	,107
Q8	-,609	,418
Q9	-,724	,414
Q10	-,791	,338

Método de extracción:  
análisis de componentes  
principales.

a. 2 componentes  
extraídos.

La matriz de varianzas-covarianzas analizada por defecto es la matriz de correlaciones entre las variables incluidas en el análisis. Puesto que esta matriz es de dimensiones  $n \times n$ , es posible extraer hasta  $n$  factores independientes. En la columna de porcentajes acumulados (*% acumulado*), se observa que con los  $n$  factores que es posible extraer se consigue explicar el 100% de la varianza total, pero con ello no se consigue el objetivo de reducir el número de dimensiones necesarias para explicar los datos.

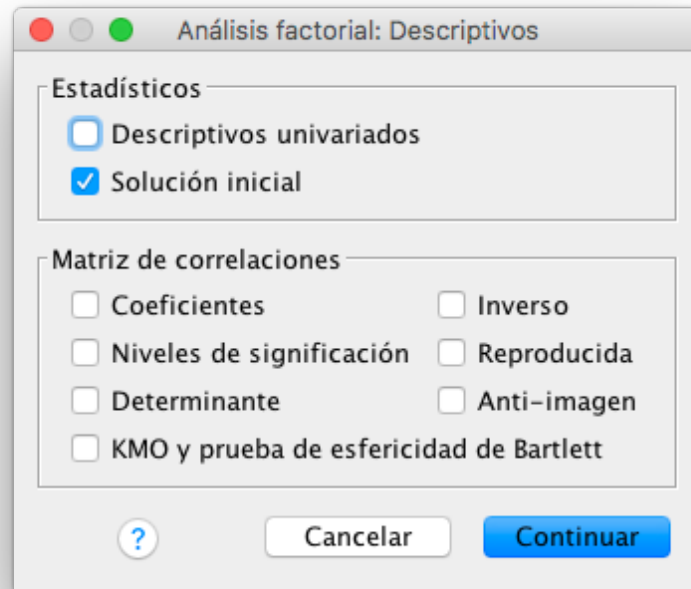
Posteriormente se muestra la *solución factorial* propiamente dicha. Contiene las correlaciones entre las variables originales (o **saturaciones**) y cada uno de los factores. Conviene señalar que esta matriz cambia de denominación dependiendo del método de extracción elegido. En este caso se denomina **matriz de componentes** porque por defecto se utiliza el método de *componentes principales* como método de extracción. En otros métodos recibe el nombre de matriz de **estructura factorial**.

Comparando las saturaciones relativas de cada variable en cada uno de los factores puede apreciarse que variables constituyen cada factor.



## 1. Descriptivos

La opción **Descriptivos** ofrece algunos estadísticos descriptivos, además de la matriz de correlaciones y otras matrices y estadísticos relacionados con ella. Para obtener estos estadísticos hay que pulsar sobre el botón Descriptivos... del cuadro de diálogo Análisis factorial:



**Estadísticos.** Este apartado contiene las opciones que permiten seleccionar los estadísticos descriptivos del análisis:

- **Descriptivos univariados.** Muestra, para cada variable, el número de casos válidos, la media y la desviación típica. Si se mantienen las especificaciones que el programa tiene establecidas por defecto y el análisis se basa en la matriz de correlaciones, las diferencias de escala y de variabilidad entre las variables carecen de relevancia. Sin embargo, si se decide que el análisis se base en la matriz de varianzas-covarianzas, las variables con mayor variabilidad tendrán mayor importancia en la solución final.

- **Solución inicial.** Permite obtener las comunalidades iniciales, los autovalores de la matriz analizada y los porcentajes de varianza asociados a cada autovalor. Esta opción actúa por defecto.

**Matriz de correlaciones.** En este apartado se encuentran las opciones necesarias para obtener información sobre la matriz de correlaciones y algunos estadísticos asociados a ella.

- **Coeficientes.** Muestra la matriz con los coeficientes de correlación entre las variables



utilizadas en el análisis, es decir, los coeficientes de correlación de Pearson entre cada par de variables. Si no se especifica lo contrario, ésta es la matriz de la cual parte el análisis. Con el método de extracción 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 de correlaciones.

Para que el análisis sea fructífero es conveniente 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.

- **Niveles de significación.** Incluye en la matriz de correlaciones los niveles críticos unilaterales asociados a cada coeficiente. Un nivel crítico menor que 0,05 indica que la correlación poblacional entre el correspondiente par de variables puede ser considerada significativamente distinta de cero. Lo deseable, por tanto, es encontrar muchos niveles críticos pequeños.

- **Determinante.** Muestra el determinante de la matriz de correlaciones. El valor del determinante aparece en una nota a pie de tabla. Los determinantes próximos a cero están indicando que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial es una técnica pertinente para analizar esas variables. No obstante, si el determinante de la matriz de correlaciones vale exactamente cero, el programa emite una advertencia indicando que no es posible calcular la matriz inversa, en cuyo caso tampoco será posible utilizar algunos de los métodos de extracción (por ejemplo, ejes principales o máxima verosimilitud).

- **Inversa.** Muestra la inversa de la matriz de correlaciones. Esta matriz es la base para el cálculo de las comunales iniciales en algunos métodos de extracción y para el cálculo de la matriz anti-imagen.

- **Reproducida.** Muestra la matriz reproducida. La matriz de correlaciones reproducidas contiene las correlaciones que es posible reproducir utilizando tan sólo la información contenida en la solución factorial. En concreto, la matriz reproducida se obtiene post-multiplicando la matriz factorial por su traspuesta. Si el modelo es bueno y el número de factores el adecuado, la estructura factorial debe ser capaz de reproducir la matriz de correlaciones. En la diagonal de la matriz reproducida se encuentran las comunales finales.

Además de la matriz de correlaciones reproducidas, también se incluye la matriz residual, la cual contiene 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. Si el análisis ha sido fructífero, la mayoría de las correlaciones reproducidas se parecerán a las correlaciones observadas y los residuos serán





muy pequeños. De hecho, como orientación, la tabla incluye una nota a pie de tabla que contabiliza el número de residuos mayores que 0,05 (un valor arbitrariamente pequeño) y el porcentaje que ese número representa sobre el total de correlaciones no redundantes de la matriz.

Existen varias razones por las que el análisis podría desembocar en una matriz residual con un gran número de residuos altos (en valor absoluto). En primer lugar, podría ocurrir que se hubiera extraído un número insuficiente de factores y que, consecuentemente, la estructura factorial no fuera capaz de reproducir adecuadamente la matriz de correlaciones observada. En segundo lugar, podría ocurrir que las correlaciones observadas estuvieran mal estimadas, bien por la presencia de sesgos en la medida de las variables, bien porque el coeficiente de correlación de Pearson no fuera el apropiado para cuantificar la relación por causa de la escala utilizada para medir las variables. Por último, podría ocurrir que el modelo factorial no fuera pertinente para analizar los datos (porque las variables no están linealmente relacionadas, porque en los datos analizados no existe ningún tipo de estructura factorial, etc.).

- **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 y la matriz de *correlaciones anti-imagen* contiene los coeficientes de correlación parcial cambiados de signo.

En este contexto, un **coeficiente de correlación parcial** expresa el grado de relación existente entre dos variables tras eliminar el efecto de las restantes variables incluidas en el análisis. Cuando las variables incluidas en el análisis comparten gran cantidad de información debido a la presencia de factores comunes, la correlación parcial entre cualquier par de variables debe ser reducida. Por el contrario, cuando dos variables comparten gran cantidad de información entre ellas, pero no la comparten con las restantes variables (ni, consecuentemente, con los factores comunes), la correlación parcial entre ellas será elevada, siendo esto un mal síntoma de cara a la idoneidad del análisis.

Por otro lado, las correlaciones parciales son también estimaciones de las correlaciones entre los factores únicos (existe un factor único para cada variable del modelo). Y puesto que los factores únicos son independientes entre sí, las correlaciones parciales deben ser próximas a cero.

La correlación anti-imagen es el negativo de la correlación parcial entre dos variables. Si la matriz de correlaciones anti-imagen contiene una gran proporción de coeficientes elevados, el modelo factorial puede no ser adecuado para analizar los datos.

La diagonal de la matriz de correlaciones anti-imagen contiene una medida de adecuación muestral para cada variable. Esta medida es similar a la medida KMO, pero para cada variable individualmente considerada.

Los valores de la diagonal de la matriz de covarianza anti-imagen se obtienen restando a 1 la



correlación múltiple al cuadrado entre cada variable y las restantes variables del análisis. Representan, por tanto, una estimación de la unicidad de cada variable, o lo que es lo mismo, una estimación de lo que cada variable tiene de propio o de no compartido con las demás. Habitualmente, los valores de estas dos matrices se muestran en notación científica (en formato exponencial).

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

• **KMO y prueba de esfericidad de Bartlett.** La medida de *adecuación muestral* KMO (Kaiser-Meyer-Olkin) contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas. Compara la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial.

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} r_{ij.m}^2}$$

donde  $r_{ij}$  representa el coeficiente de correlación simple entre las variables  $i$  y  $j$  y  $r_{ij.m}$  representa la correlación parcial entre las variables  $i$  y  $j$  eliminado el efecto de las restantes  $m$  variables incluidas en el análisis.

El estadístico KMO varía entre 0 y 1.

Puesto que la correlación parcial entre dos variables debe ser pequeña cuando el modelo factorial es adecuado, el denominador debe aumentar poco si los datos corresponden a una estructura factorial, en cuyo caso KMO tomará un valor próximo a 1.

Si el valor de la medida de adecuación muestral es reducido (los valores por debajo de 0,6 se consideran mediocres) puede que no sea pertinente utilizar el análisis factorial con esos datos. Los menores que 0,5 indican que no debe utilizarse el análisis factorial con los datos muestrales que se están analizando. (La diagonal de la matriz de correlaciones anti-imagen incluye los coeficientes de adecuación muestral para cada variable individualmente considerada).

Por su lado, la *prueba de esfericidad de Bartlett* contrasta la hipótesis nula de que la matriz de correlaciones observada es en realidad una matriz identidad. Asumiendo que los datos provienen de una distribución normal multivariante, el estadístico de Bartlett se distribuye aproximadamente según el modelo de probabilidad *chi-cuadrado* y es una transformación del determinante de la matriz de correlaciones. Si el nivel crítico (Sig.) es mayor que 0,05, no podremos rechazar la hipótesis nula de esfericidad y, consecuentemente, no podremos asegurar que el modelo factorial sea adecuado para explicar los datos.



### III. Extracción

La opción **Extracción** permite controlar varios aspectos relacionados con la fase de extracción de los factores. Entre otras cosas, permite decidir qué modelo factorial se desea utilizar, es decir, elegir la matriz de datos en la que se basará el análisis, así como cuántos factores deben extraerse. Para controlar los aspectos relacionados con el proceso de extracción de factores hay que pulsar en el botón **Extracción...** del cuadro de diálogo *Análisis factorial*.

Análisis factorial: Extracción

Método: Componentes principales

**Analizar**

☐ Matriz de correlaciones

☒ Matriz de covarianzas

**Mostrar**

☒ Solución factorial sin rotar

☐ Gráfico de sedimentación

**Extraer**

☒ Basado en autovalor

Autovalores mayores que: 0,7 veces el autovalor de la...

☐ Número fijo de factores

Factores que extraer:

N.º máximo de iteraciones para convergencia: 25

? Cancelar Continuar

**Método.** En esta lista desplegable se puede seleccionar el modelo factorial que será utilizado para estimar las saturaciones de las variables en los factores. Los distintos métodos difieren tanto en el algoritmo de cálculo como en la matriz que será analizada (se asume que la matriz seleccionada es la matriz de correlaciones). Los distintos métodos disponibles son:

- **Componentes principales.** Método de extracción en el que los factores obtenidos son los autovectores de la matriz de correlaciones re-escalados.



- **Mínimos cuadrados no ponderados.** Método de extracción que minimiza la suma de los cuadrados de las diferencias entre las matrices de correlaciones observada y reproducida, ignorando los elementos de la diagonal.
- **Mínimos cuadrados generalizados.** Método de extracción que minimiza la suma de los cuadrados de las diferencias entre las matrices de correlaciones observada y reproducida. Las correlaciones se ponderan por el inverso de su unicidad, de manera que las variables cuya unicidad es alta reciben un peso menor que aquellas cuyo valor es bajo. Este método genera un estadístico de bondad de ajuste *chi-cuadrado* que permite contrastar la hipótesis nula de que la matriz residual es una matriz nula.
- **Máxima verosimilitud.** Método de extracción que proporciona 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. Las correlaciones se ponderan por el inverso de la unicidad de las variables y se emplea un algoritmo iterativo. Este método genera un estadístico de bondad de ajuste *chi-cuadrado* que permite contrastar la bondad del modelo para explicar la matriz de correlaciones.
- **Ejes principales.** Método de estimación iterativo en el que, como estimación inicial de la comunalidad, la matriz de correlaciones original se reduce sustituyendo los unos de su diagonal por las estimaciones de la correlación múltiple al cuadrado entre cada variable y todas las demás. La matriz reducida se auto-descompone y se corrigen las estimaciones iniciales de la comunalidad por las nuevas estimaciones resultantes. El proceso continua hasta que no existe diferencia entre las estimaciones de las comunalidades entre dos pasos sucesivos o se alcanza alguno de los criterios de parada.
- **Alfa.** Método de extracción que considera las variables incluidas en el análisis como una muestra del universo de las variables posibles. Este método maximiza la generalizabilidad de los factores calculada como el *alfa de Cronbach*.
- **Imagen.** Método de extracción en el que se auto-descompone la matriz de correlaciones imagen. Se asume que la comunalidad es igual al cuadrado de la correlación múltiple entre una variable y todas las demás. Al solicitar este método de extracción, los resultados incluyen una tabla con la matriz de covarianza imagen.

**Analizar.** Las opciones de este apartado sólo están disponibles cuando se seleccionan los métodos de componentes principales, ejes principales y análisis imagen. Permite seleccionar el tipo de matriz que será analizada.

- **Matriz de correlaciones.** El análisis se basa en la matriz de correlaciones, en la matriz de correlaciones reducida, o en la matriz de correlaciones anti-imagen, según el método seleccionado.
- **Matriz de covarianza.** El análisis se basa en la matriz de varianzas-covarianzas, en la matriz de varianzas-covarianzas reducida, o la matriz de covarianzas anti-imagen, según el método seleccionado.



**Extraer.** Este apartado contiene opciones que permiten determinar el número de factores que se extraerán en la solución factorial, bien a partir de una regla heurística, bien especificando un número concreto:

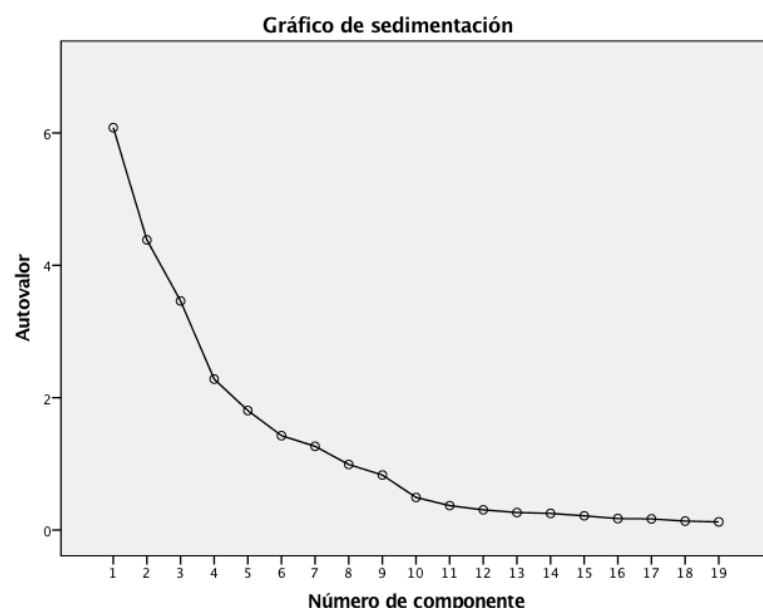
- **Autovalores mayores que.** Si la matriz analizada es la de *correlaciones*, esta opción permite utilizar el tamaño de los autovalores como criterio para decidir el número de factores que estarán presentes en la solución factorial. Por defecto se extraen los factores cuyos autovalores son mayores que la unidad (a este criterio se le denomina *regla K1*). Si la matriz analizada es la de *varianzas-covarianzas*, la regla expresa el número de veces que un autovalor debe ser mayor que el autovalor promedio de la matriz para que el correspondiente factor sea retenido en la solución.

El valor que actúa por defecto es 1, pero este valor puede cambiarse introduciendo otro distinto (entre cero y el número de variables) en el correspondiente cuadro de texto.

- **Número de factores.** Permite especificar el número exacto de factores que se desea incluir en la solución. Se debe introducir dicho número en el recuadro de texto.

**Mostrar.** Estas opciones permiten seleccionar los resultados de la extracción que aparecerán en el visor de resultados.

- **Solución factorial sin rotar.** Muestra las saturaciones factoriales sin rotar (la matriz de componentes o factorial), las comunalidades y los autovalores de la solución factorial.
- **Gráfico de sedimentación** (también llamado *prueba de sedimentación de Cattell*). Muestra una representación gráfica de la magnitud de los autovalores. Si un autovalor se aproxima a cero, esto significa que el factor correspondiente a ese autovalor es incapaz de explicar una cantidad relevante de la varianza total. Por tanto, un factor al que corresponde un autovalor próximo a cero se considera un factor residual y carente de sentido en el análisis.





Al representar todos los autovalores según su tamaño, es posible formarse muy rápidamente una idea sobre si la cantidad de varianza asociada a cada uno de ellos es relevante para el análisis o si por el contrario se trata sólo de varianza residual. Los autovalores residuales se encuentran en la parte derecha del gráfico, formando una planicie de poca inclinación, frente a la fuerte pendiente formada por los autovalores que explican la mayor parte de la varianza disponible. Por ello, es conveniente inspeccionar el gráfico de sedimentación de izquierda a derecha, buscando el punto de inflexión en el que los autovalores dejan de formar una pendiente significativa y comienzan a describir una caída de poca inclinación. El corte en la tendencia descendente sirve de regla para la determinación del número óptimo de factores que deben estar presentes en la solución.

Es importante resaltar que el gráfico de sedimentación no varía con el número de factores seleccionado. Por otra parte, el gráfico siempre muestra todos los posibles autovalores de la matriz de correlaciones original y no los autovalores de la matriz analizada, que puede ser distinta de la de correlaciones, según el método de extracción utilizado.

**Nº de iteraciones para convergencia.** Este cuadro de texto permite establecer el número máximo de iteraciones que los algoritmos pueden realizar para encontrar la solución factorial final. El valor por defecto es 25, habitualmente suficiente para obtener una solución. Este valor puede cambiarse introduciendo un entero positivo.



## IV. Rotación

La opción **Rotación** permite controlar la fase de rotación del análisis. Con esta opción puede definirse el método de rotación que deseamos utilizar para facilitar la interpretación de la solución factorial y solicitar la representación gráfica de las saturaciones. Por defecto, no se encuentra seleccionado ningún método de rotación. Para seleccionar el método de rotación hay que pulsar en el botón **Rotación...** del cuadro de diálogo *Análisis factorial*.

Análisis factorial: Rotación

Método

☐ Ninguno ☐ Quartimax

☐ Varimax ☐ Equamax

☒ Oblimin directo ☐ Promax

Delta:  Kappa:

Mostrar

☒ Solución rotada ☒ Gráficos de cargas

N.º máximo de iteraciones para convergencia:

? Cancelar Continuar

La rotación de la solución original se realiza con el objetivo de mejorar la interpretación de la estructura factorial. Las restricciones de la auto-descomposición de la matriz de correlaciones imponen que el primer factor explique el máximo de la varianza común disponible en los datos, que el segundo factor explique el máximo de la varianza común restante (e independiente de la explicada por el primer factor), y así sucesivamente hasta el último de los factores. Estas restricciones se imponen para deshacer la indeterminación intrínseca a la solución del sistema homogéneo de ecuaciones que da lugar a los autovectores. Un efecto indeseable de estas restricciones es que los primeros factores tienden a capitalizar la información de covariación contenida en la matriz de



correlaciones, acumulando más información de la que posiblemente les corresponda. Este hecho se aprecia en que las saturaciones de las variables en los primeros factores (y en especial en el primer factor) suelen encontrarse infladas, llevando esto a conceder excesiva importancia a los primeros factores. Cuando la estructura factorial es clara y cada variable del análisis se encuentra inequívocamente asignada a un único factor, el efecto “contaminante” de las restricciones no suele apreciarse. Sin embargo, cuando las variables saturan en más de un factor o existe un factor general que domina la solución, la rotación puede ser de gran utilidad para interpretar los resultados.

Otro de los motivos que justifican la rotación es que la solución factorial original es siempre ortogonal (los factores no rotados son siempre independientes entre sí). Sin embargo, existe un gran número de situaciones (y en especial en las ciencias sociales) en las que los factores pueden estar relacionados entre sí. En estos casos, si se desea estimar el grado de relación existente entre los factores, debe recurrirse a una rotación oblicua.

Cuando se aplica una rotación, la tabla de comunalidades es la misma que la obtenida en la extracción no rotada. Es importante resaltar este aspecto, pues el proceso de rotación busca clarificar la interpretación de la estructura factorial sin alterar la situación relativa de unas variables respecto a las otras, y sin alterar tampoco el porcentaje de la varianza de cada variable que es capaz de explicar cada factor. Para que la comunalidad de las variables cambie es necesario variar el número de factores de la solución.

Lo que sí cambia en el proceso de rotación es el porcentaje de varianza total explicada por cada factor (y cambia tanto más cuanto más éxito tiene la rotación). Al rotar la solución, la tabla de porcentajes de varianza explicada incorpora información adicional referente a la suma de las saturaciones tras la rotación de los factores. Si las sumas de los cuadrados de las saturaciones difieren poco, puede concluirse que la rotación utilizada no mejora demasiado la interpretación de la solución.

En definitiva, el proceso de rotación busca lo que Thurstone (1947) denominó una *estructura simple*: variables que saturan, a ser posible, en un único factor, y factores que contengan un número reducido de variables que saturan inequívoca y exclusivamente en ellos. Con todo, las variables que compartan información con varios factores, si existen, entorpecerán el proceso de rotación y, en lugar de una única saturación elevada en un único factor, tenderán a mostrar saturaciones moderadas en varios factores.

Al aplicar una rotación aparece la matriz de *transformación de los factores*, que es la matriz utilizada para rotar la solución inicial. Esta matriz adopta la forma:

$$\Lambda^* = \Lambda T$$

$$\text{con } \begin{cases} T = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \text{ si la rotación se hace en el sentido de las agujas del reloj} \\ T = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \text{ si la rotación se hace en el sentido contrario al de las agujas del reloj} \end{cases}$$





Donde  $\Lambda$  es la matriz de estructura factorial antes de la rotación,  $T$  es la matriz de transformación y  $\Lambda^*$  es la matriz de estructura factorial después de la rotación.

**Método.** En este apartado se puede seleccionar el método de rotación de la solución factorial. Se encuentran disponibles tres procedimientos de rotación *ortogonal*, mediante los cuales se respeta la independencia entre factores de la solución inicial, y dos procedimientos de rotación *oblicua*, mediante los cuales pueden obtenerse factores relacionados entre sí. Los métodos disponibles son:

- **Ninguno.** No se aplica ningún método de rotación. Es la opción que actúa por defecto. Cuando la solución consta de un único factor y no se ha marcado esta opción, el Visor de resultados muestra un mensaje de advertencia.
- **Varimax.** Método de rotación ortogonal que minimiza el número de variables que tienen saturaciones altas en cada factor. Simplifica la interpretación de los factores optimizando la solución por columna.
- **Quartimax.** Método de rotación ortogonal que minimiza el número de factores necesarios para explicar cada variable. Simplifica la interpretación de las variables observadas optimizando la interpretación por filas.
- **Equamax.** Método de rotación que es combinación del método Varimax, que simplifica los factores, y el método Quartimax, que simplifica las variables. Se minimiza tanto el número de variables que saturan alto en un factor como el número de factores necesarios para explicar una variable.
- **Oblimin directo.** Método para la rotación oblicua (no ortogonal). Cuando delta es igual a cero (el 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, puede introducirse un número menor o igual que 0,8.

**Delta.** El valor de delta permite controlar el grado de oblicuidad que pueden llegar a alcanzar los factores de la solución.

- **Promax.** Rotación oblicua que permite que los factores estén correlacionados. Puede calcularse más rápidamente que una rotación Oblimin directa, por lo que es útil para grandes conjuntos de datos.

**Kappa.** Parámetro que controla el cálculo de la rotación Promax. El valor por defecto es 4. Este valor es adecuado para la mayoría de los análisis.

**Mostrar.** Este cuadro se permite seleccionar los resultados de la rotación que se mostrarán en el *Visor de resultados*. Por defecto se muestra la solución rotada cuando se selecciona alguno de los métodos de rotación. Si se encuentra seleccionada la opción *Ninguna* del recuadro Método no será posible seleccionar ninguna de las opciones de este recuadro.



- **Solución rotada.** Permite obtener una o más tablas con los resultados del proceso de rotación. Al seleccionar una rotación ortogonal, esta opción permite obtener la matriz de estructura factorial rotada y la matriz de transformación necesaria para rotar los factores a partir de la solución inicial. Además, en la tabla de porcentajes de varianza explicada aparecen columnas adicionales que contienen la varianza total explicada por los factores rotados. Al seleccionar una rotación oblicua, esta opción permite obtener la matriz de configuración rotada, que contiene las saturaciones de las variables en los factores, y la matriz de estructura, que contiene las correlaciones entre las variables observadas y los factores (cuando la rotación es ortogonal, ambas matrices son idénticas). Además, ofrece la matriz de correlaciones entre los factores y desecha la matriz de transformación para la rotación. En la tabla de *porcentajes de varianza* explicada sólo se incluyen los autovalores de los factores rotados (ya que no tiene sentido hablar de porcentajes de varianza independientes).

- **Gráficos de saturaciones.** Representa el espacio factorial definido por los factores contenidos en la solución factorial. Si la solución contiene un único factor, el gráfico no se genera y aparece una advertencia indicando tal circunstancia; si la solución contiene dos factores se genera un diagrama de dispersión simple; si la solución contiene tres o más factores se genera un gráfico de dispersión tridimensional en el que sólo se representan los tres primeros factores. Cuando la solución contiene más de tres factores, el gráfico tridimensional representa los tres primeros factores, pero almacena también la información correspondiente a los restantes factores. Para representar factores distintos de los tres primeros pueden hacerse dos cosas, ambas desde el Editor de gráficos (al cual se accede pinchando dos veces sobre un gráfico): 1) seleccionar en el menú Series los tres factores que se desea representar, o 2) solicitar en el menú Galería un diagrama de dispersión matricial para representar simultáneamente todos los factores dos a dos.

Un gráfico de saturaciones factoriales es un diagrama de dispersión en el que los factores definen los ejes del espacio y las variables constituyen los puntos del diagrama. Las coordenadas de una variable en cada factor se corresponden con las saturaciones de la variable en dichos factores, es decir, con los valores de la matriz factorial.

**Nº máximo de iteraciones para convergencia.** Permite determinar el número máximo de iteraciones que puede recorrer el algoritmo para la estimación de la solución rotada. Por defecto se efectúan un máximo de 25 iteraciones, lo que es suficiente para la mayoría de las situaciones.

## **Diferencias entre rotaciones ortogonales y oblicuas**

A diferencia de lo que ocurre en la rotación ortogonal, los resultados de la rotación oblicua no pueden representarse en una única matriz. Si los factores son ortogonales (independientes entre sí), la saturación de una variable en un factor, es decir, su proyección sobre el factor, es igual a la correlación de esa variable con el factor. Pero si los factores son oblicuos (correlacionan entre sí), la saturación y



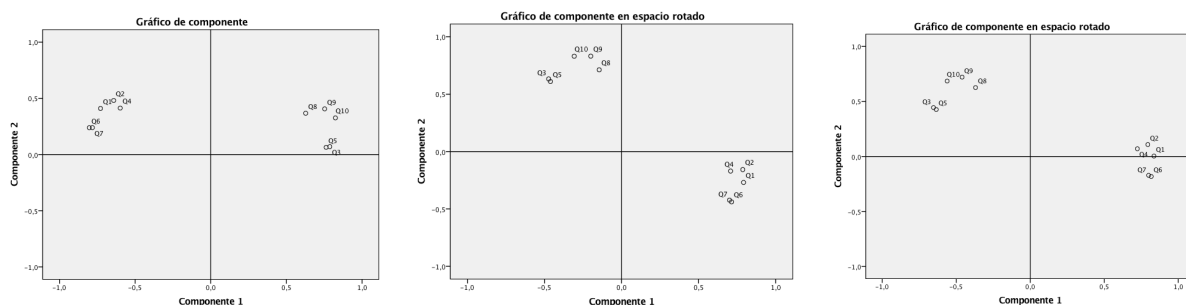
la correlación de una variable en un factor no coinciden. Por este motivo, al solicitar una rotación oblicua, el *Visor* muestra dos matrices para la estructura factorial rotada: una con las correlaciones (a la que llama *matriz factorial*) y otra con las saturaciones (a la que llama *matriz de configuración*).

La *matriz de configuración* ofrece las saturaciones de las variables en los factores de la solución rotada. Esas saturaciones, que son las que se representan en el gráfico del espacio factorial rotado, representan la contribución neta de cada variable en cada factor, por lo que constituyen la manera más fácil de interpretar la solución factorial.

La *matriz de estructura* contiene las correlaciones de las variables con los factores de la solución rotada. Estas correlaciones representan la contribución bruta de cada variable a cada factor. Cuando los factores correlacionan mucho entre sí (se encuentran muy próximos en el espacio), la matriz de estructura contiene correlaciones muy grandes entre todas las variables y todos los factores, lo cual hace muy difícil la interpretación por la imposibilidad de precisar a qué factor único hay asignar cada variable (si bien esto representa la situación real: las variables que correlacionan con un factor también lo harán con los factores relacionados con él).

Puede darse el caso de que, aun solicitando una rotación oblicua, los factores permanezcan ortogonales. El algoritmo de rotación oblicua busca rotar de manera autónoma cada uno de los factores, pero eso no quiere decir que por ello los factores deban aproximarse entre sí cuando la solución ortogonal es la mejor de las posibles

Cuando se realiza una rotación oblicua, el gráfico de las *saturaciones en el espacio factorial rotado* puede resultar engañoso. Aunque el gráfico representa la posición relativa de las variables en los factores, el ángulo entre los factores se mantiene en 90 grados, independientemente del ángulo real obtenido con la rotación. Para interpretar correctamente el gráfico debe tenerse en cuenta cuál es la posición de las variables respecto del factor en el que más saturan (que es el factor al que se encuentran más próximas). Sin embargo, la inclinación de los ejes debe intuirse a partir de los valores de la *matriz de correlaciones entre los factores*.





## V. Puntuaciones factoriales

Una vez alcanzada la solución factorial final, suele resultar interesante obtener una estimación de las puntuaciones de los sujetos en cada uno de los factores resultantes de la extracción a fin de valorar la situación relativa de cada sujeto en esas “dimensiones ocultas” capaces de resumir la información contenida en las variables originales. El cuadro de diálogo *Puntuaciones factoriales* contiene las opciones que permiten solicitar las estimaciones de las *puntuaciones factoriales* y seleccionar el método de estimación que se desea utilizar para obtener tales estimaciones. Para acceder a estas opciones hay que pulsar en el botón **Puntuaciones...** del cuadro de diálogo *Análisis factorial*.



Además SPSS muestra la opción **Guardar como variables**. Activando esta opción se guardan automáticamente en el *Editor de datos* las puntuaciones factoriales estimadas para cada sujeto en cada uno de los factores obtenidos en la solución factorial. Para ello, el SPSS crea en el archivo de datos activo tantas variables nuevas como factores contenga la solución factorial. Si no se selecciona esta opción no es posible acceder a los métodos de estimación de las puntuaciones factoriales.

**Método.** Este apartado contiene varios métodos de estimación de las puntuaciones factoriales. Por defecto se encuentra seleccionado el método de *Regresión*, que es el de uso más generalizado. Es importante señalar que las opciones de este recuadro no tienen efecto alguno cuando se ha seleccionado *componentes principales* como método de extracción, ya que en ese modelo factorial las puntuaciones factoriales no son estimadas, sino calculadas directamente a partir de las variables originales.



- **Regresión.** Método de estimación de las puntuaciones factoriales en el que las estimaciones resultantes tienen una media de cero y una varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Las puntuaciones factoriales estimadas con este método pueden estar correlacionadas incluso cuando los factores son ortogonales.
- **Bartlett.** Método de estimación de las puntuaciones factoriales en el que las estimaciones resultantes tienen una media de cero. Este método 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.** Este método de estimación es una modificación del método de Bartlett que asegura la ortogonalidad de las puntuaciones factoriales estimadas. Las estimaciones resultantes tienen una media de cero, una desviación típica de uno y son independientes entre sí (incluso en el caso de que se haya solicitado una solución rotada oblicua).
- **Mostrar matriz de coeficientes de las puntuaciones factoriales.** Esta opción permite obtener una tabla con los pesos o ponderaciones necesarios para calcular las puntuaciones factoriales a partir de las variables originales. Esta opción se encuentra desactivada por defecto. Por tanto, para obtener la matriz de coeficientes no basta con solicitar las puntuaciones factoriales.

Combinando cada variable con sus correspondientes coeficientes pueden construirse las dos ecuaciones lineales en las que se basa el cálculo de las puntuaciones factoriales:

Las dos puntuaciones factoriales de un sujeto se obtienen sustituyendo cada variable por sus respectivos valores.

Las puntuaciones factoriales se encuentran en formato *diferencial*, por lo que una puntuación de cero 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. Si se desea eliminar los signos negativos siempre es posible realizar una transformación de las puntuaciones para cambiar la escala de las nuevas variables.

Las puntuaciones factoriales se almacenan de manera automática en el *Editor de datos* y reciben de forma automática un nombre que identifica, por este orden, el método de estimación de las puntuaciones (en el ejemplo, *REGR*), el número del factor al que corresponden las puntuaciones (*factor score 1*) y el número del análisis (*analisis 1*). Este nombre es único y distintivo, de manera que si se solicitan nuevas estimaciones de las puntuaciones, las nuevas puntuaciones se almacenarán al final del archivo de datos con nuevos nombres.

Para interpretar mejor las puntuaciones factoriales de los sujetos conviene solicitar algunos estadísticos descriptivos de las nuevas variables mediante el procedimiento **Estadísticos descriptivos > Descriptivos** del menú **Analizar**.

Cuando la extracción de los factores se realiza con el método *componentes principales*, las



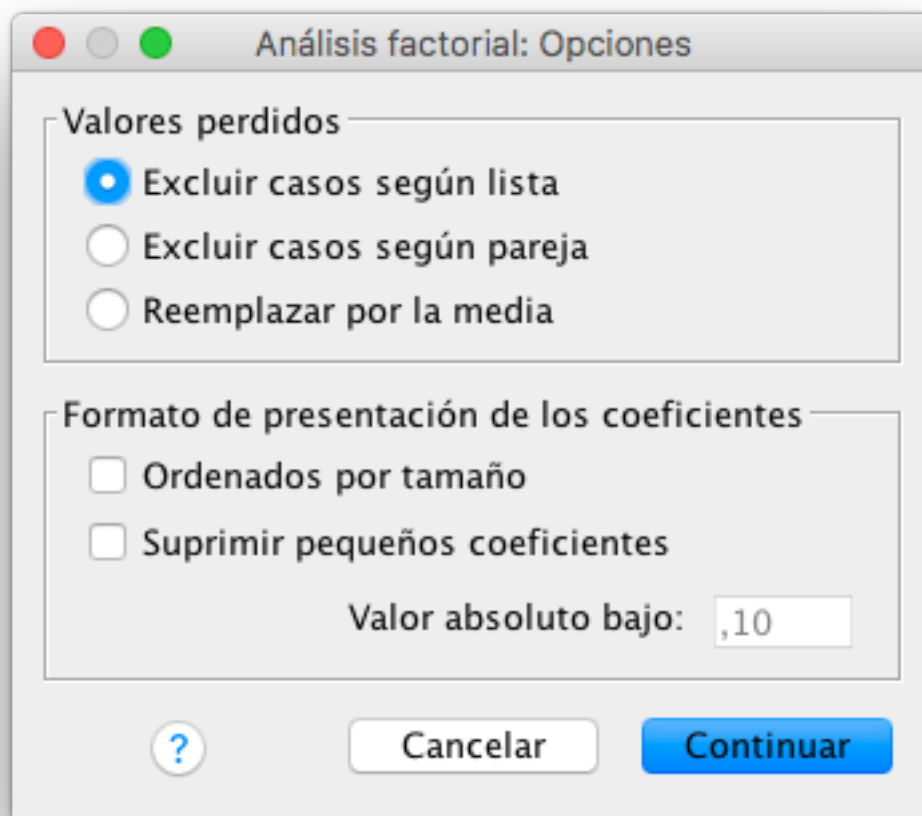
puntuaciones factoriales no se obtienen mediante estimación, sino que son directamente calculadas a partir de la solución factorial. Y puesto que la extracción con el método *componentes principales* siempre ofrece una solución ortogonal, las puntuaciones factoriales basadas en esa solución también serán ortogonales.

Sin embargo, cuando se utiliza un método de extracción distinto del de *componentes principales* no es posible obtener directamente las puntuaciones factoriales a partir de la matriz de estructura, sino que deben ser estimadas mediante uno cualquiera de los métodos de estimación disponibles.



## VI. Opciones

SPSS permite controlar algunos aspectos relacionados con el tratamiento que deben recibir los valores perdidos y el formato de las tablas de resultados que genera el *Visor de resultados*. Para controlar estos aspectos es necesario pulsar en el botón Opciones... del cuadro de diálogo *Análisis factorial*.



**Valores perdidos.** Este recuadro permite controlar el tratamiento que se desea dar a valores perdidos.

- **Excluir casos según lista.** Es la opción por defecto. Se excluyen del análisis los sujetos que tengan valores perdidos en cualquiera de las variables trasladadas a la lista Variables del cuadro de diálogo *Análisis factorial*. Es el tratamiento más consistente de todos: sólo se incluyen en el análisis los casos *completos* (es decir, los casos con puntuación válida en todas las variables seleccionadas). Sin embargo, conviene tener en cuenta que esta forma de tratar



los valores perdidos puede suponer la pérdida de un gran número de casos y la consiguiente reducción del tamaño efectivo de la muestra.

- **Excluir casos según pareja.** Los sujetos con valor perdido en una variable se excluyen del análisis sólo para el cálculo de los estadísticos en los que esté implicada esa variable. Este método permite aprovechar más cantidad de información que el anterior, pero, puesto que no todas las correlaciones se calculan sobre el mismo número de sujetos, podrían obtenerse matrices de correlaciones inconsistentes imposibles de analizar posteriormente.
- **Reemplazar por la media.** Los valores perdidos en una variable se sustituyen por la media de esa variable. Si en una variable existen muy pocos casos con valor perdido, reemplazar el valor perdido por la media no constituye un problema importante. Pero en la medida en que el número de valores perdidos aumenta, la sustitución por la media tiene el efecto de *centrar* las variables disminuyendo su variabilidad.

**Formato de visualización de los coeficientes.** Las opciones de este apartado permiten cambiar algunos aspectos relacionados con el formato de presentación de las tablas.

- **Ordenados por tamaño.** Esta opción sirve para ordenar las variables de las tablas de resultados en función de la magnitud (en valor absoluto) de los coeficientes de esas tablas (saturaciones, correlaciones, etc). La ordenación se realiza de forma ascendente: primero las variables con coeficientes más altos. Si no se marca esta opción, las tablas muestran las variables en el mismo orden en el que han sido trasladadas a la lista de Variables del cuadro de diálogo *Análisis factorial*.
- **Suprimir valores absolutos menores que....** Esta opción permite suprimir de las tablas de resultados los coeficientes cuyo valor absoluto sea menor que el valor establecido en el cuadro de texto. El valor por defecto es 0,10, pero este valor puede cambiarse introduciendo un valor distinto. Esta opción es de gran ayuda: al desaparecer de la tabla los coeficientes excesivamente pequeños (en valor absoluto), se facilita notablemente la interpretación de los resultados.