Análisis de Componentes Principales

Dr. José Dionicio Zacarias Flores

Introducción

- Un análisis de componentes principales concierne a explicar la estructura varianza-covarianza de un conjunto de variables a través de pocas combinaciones lineales de esas variables.
- Sus objetivos generales son:

  a) reducción de datos

  b) interpretación

Introducción

- Aunque p variables son requeridas para reproducir al sistema total de esta variabilidad, después de un cierto trabajo este puede ser explicado por un número pequeño k de ***componentes principales***.
- Es decir, las k componentes contienen casi la misma información que las p variables originales, por lo tanto las k componentes pueden reemplazar a las p variables originales.

Introducción

- Un análisis de componentes principales a menudo revela una relación que no se sospechaba anteriormente y, por lo tanto, permite interpretaciones que normalmente no resultarían.
- El análisis de los componentes principales es más un medio para un fin en lugar de un fin en sí mismo, porque con frecuencia sirven como pasos intermedios en investigaciones mucho más grandes.

Introducción

- Por ejemplo, pueden ser la entrada a una regresión múltiple o análisis de cluster. Más aun, componentes principales son una "factorización" de la matriz de covarianza para el modelo de análisis de factores.

Componentes principales de la población

- Algebraicamente, los componentes principales son combinaciones lineales particulares de las p variables aleatorias X_1, X_2, ..., X_p. Geométricamente, estas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenido al rotar el sistema original con X_1, X_2, ..., X_p como los ejes de coordenadas. Los nuevos ejes representan las direcciones con máxima variabilidad y proporciona una descripción más simple y parsimoniosa de la estructura de covarianza.

- Sea el vector aleatorio X' = {X_1, X_2, …, X_p} que tiene matriz de covarianza Σ con valores propios $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.
- Considere las combinaciones lineales

$$\begin{aligned} Y_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ Y_2 &= \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ &\vdots \qquad\qquad\qquad \vdots \\ Y_p &= \mathbf{a}_p'\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned}$$

$$\begin{aligned} \operatorname{Var}(Y_i) &= \mathbf{a}_i'\Sigma\mathbf{a}_i \qquad i = 1, 2, \ldots, p \\ \operatorname{Cov}(Y_i, Y_k) &= \mathbf{a}_i'\Sigma\mathbf{a}_k \qquad i, k = 1, 2, \ldots, p \end{aligned}$$

- Las componentes principales son esas combinaciones lineales no correlacionadas Y_1, Y_2. …, Y_p cuyas varianzas son tan grandes como sea posible.

Determinación de la primera componente principal

- La primera componente principal es la combinación lineal con varianza mínima. Es decir, maximizar $Var(Y_i) = a_i' \sum a_i$, i = 1, 2, …, p, es claro que $Var(Y_1) = a_1' \sum a_1$ puede incrementarse multiplicando a todo a_1 por cualquier constante. Para eliminar esta indeterminación, es conveniente restringir la atención a vectores con coeficientes de longitud uno.

Determinación de las componentes principales

- Así, podemos definir:
- Primera componente principal = combinación lineal $a_1'X$ que maximiza Var$(a_1'X)$ sujeta a $a_1'a_1 = 1$.
- Segunda componente principal = combinación lineal $a_2'X$ que maximiza Var$(a_2'X)$

  sujeta a $a_2'a_2 = 1$ y Cov$(a_1'X, a_2'X)$ = 0.
- En el i-ésimo paso

i-ésima componente principal = combinación lineal $a_i'X$ que maximiza Var$(a_i'X)$ sujeta a $a_i'a_i = 1$ y Cov$(a_i'X, a_k'X)$ = 0 para k < i

Teorema 1

- Sea Σ la matriz de covarianza asociada con el vector aleatorio $X' = [X_1, X_2, \ldots, X_p]$. Sea Σ que tiene a las parejas de valores y vectores propios (λ_1,e_1), (λ_2,e_2), …, (λ_p,e_p) donde $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Entonces, la i-ésima componente principal es dada por $Y_i = e_i' X = e_{i1}X_1 + e_{i2}X_2 + \quad + e_{i1}X_1$.
- Con esas elecciones,

$$Var(Y_i) = e_i' \sum e_i = \lambda_i \qquad i = 1, 2, \ldots, p$$

$$Cov(Y_i, Y_k) = e_i' \sum e_k = 0 \qquad i \neq k$$

Si algunos λ_i son iguales, las elecciones de los correspondientes vectores de coeficientes e_i, y de aquí Y_i, no son únicos.

Teorema 2

- Sea X' = $[X_1, X_2, \ldots, X_p]$ teniendo matriz de covarianza Σ, con parejas de valores y vectores propios $(\lambda_1,e_1), (\lambda_2,e_2), \ldots, (\lambda_p,e_p)$ donde $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Sean $Y_1 = e_1' X$, $Y_2 = e_2' X$, …, $Y_p = e_p' X$ son las componentes principales. Entonces

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

- El teorema 2 nos dice que:

la varianza poblacional total

$$= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$$
$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

Consecuentemente, la proporción de varianza total debido a la k-ésima componente principal es

$$= \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \qquad k = 1, 2, \ldots, p$$

Si la mayoría (por decir, de 80 a 90%) de la varianza poblacional total, para p grande, puede ser atribuible a la primera, segunda, o tercera componentes, entonces esas componentes pueden ***reemplazar*** a las p variables originales sin mucha pérdida de información.

- Cada componente del vector de coeficientes $e_i' = \left[e_{i1},\ \ldots, e_{ik}, \ldots, e_{ip}\right]$ ameritan una inspección.
- La magnitud de e_{ik} mide la importancia de la k-ésima variable a la i-ésima componente principal, codependiente de las demás variables. En particular, e_{ik} es proporcional al coeficiente de correlación entre Y_i y X_k.

Teorema 3

- Si $Y_1 = e_1' X$, $Y_2 = e_2' X$, …, $Y_p = e_p' X$ son las componentes principales obtenidas desde la matriz de covarianzas Σ, entonces

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i, k = 1, 2, \ldots, p$$

son los coeficientes de correlación entre las componentes Y_i y las variables X_k. Aquí (λ_1, e_1), (λ_2, e_2), …, (λ_p, e_p) son las parejas de valores y vectores propios para Σ.

Recomendaciones

- Los estadísticos recomiendan que solo los coeficientes e_{ik}, y no las correlaciones sean usadas para la interpretación de las componentes.
- Aunque los coeficientes y las correlaciones pueden conducir a diferentes posiciones cuando medimos la importancia de las variables a una componente dada, es nuestra experiencia de que esas posiciones a menudo no se aprecian diferencias.

Recomendaciones

- En la práctica, variables con coeficientes relativamente grandes (en valor absoluto) tienden a tener coeficientes relativamente grandes, así las dos medidas de importancia, la primera multivariable y la segunda univariada, frecuentemente dan resultados similares.
- Recomendamos que ambos, los coeficientes y correlaciones son examinados para ayudar a interpretar las componentes principales.

Ejemplo: calculando las componentes principales de la población

- Supongamos que las variables aleatorias X_1, X_2 y X_3 tienen la matriz de covarianza

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- Puede verificarse que los pares de valores y vectores propios son

$$\begin{aligned} \lambda_1 &= 5.83, & \mathbf{e}_1' &= [.383, -.924, 0] \\ \lambda_2 &= 2.00, & \mathbf{e}_2' &= [0, 0, 1] \\ \lambda_3 &= 0.17, & \mathbf{e}_3' &= [.924, .383, 0] \end{aligned}$$

Ejemplo: continuación

- Por lo tanto, las componentes principales serán:

$$
\begin{aligned}
Y_1 &= \mathbf{e}_1'\mathbf{X} = .383X_1 - .924X_2 \\
Y_2 &= \mathbf{e}_2'\mathbf{X} = X_3 \\
Y_3 &= \mathbf{e}_3'\mathbf{X} = .924X_1 + .383X_2
\end{aligned}
$$

- La variable X_3 es una de las componentes principales, porque no está correlacionada con las otras dos variables. Podemos también verificar que:

Ejemplo: continuación

- Podemos también verificar que:

$$
\begin{aligned}
\mathrm{Var}(Y_1) &= \mathrm{Var}(.383X_1 - .924X_2) \\
&= (.383)^2\,\mathrm{Var}(X_1) + (-.924)^2\,\mathrm{Var}(X_2) \\
&\quad + 2(.383)(-.924)\,\mathrm{Cov}(X_1, X_2) \\
&= .147(1) + .854(5) - .708(-2) \\
&= 5.83 = \lambda_1 \\
\mathrm{Cov}(Y_1, Y_2) &= \mathrm{Cov}(.383X_1 - .924X_2, X_3) \\
&= .383\,\mathrm{Cov}(X_1, X_3) - .924\,\mathrm{Cov}(X_2, X_3) \\
&= .383(0) - .924(0) = 0
\end{aligned}
$$

- Además la varianza poblacional total es:

$$\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 5 + 2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2.00 + .17$$

Ejemplo: continuación

- La proporción de varianza total a considerar por parte de la primera componente principal es:

$$\frac{\lambda_1}{(\lambda_1 + \lambda_2 + \lambda_3)} = \frac{5.83}{8} = .73$$

- Además, los dos primeros componentes representan una proporción (5.83 + 2)/8 = .98 de la varianza poblacional. En este caso, las componentes Y_1 y Y_2 reemplazan a las tres variables originales con poca pérdida de información.

Ejemplo: continuación

- Utilizando el teorema 3 se obtiene:

$$\rho_{Y_1,X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{.383\sqrt{5.83}}{\sqrt{1}} = .925$$

$$\rho_{Y_1,X_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-.924\sqrt{5.83}}{\sqrt{5}} = -.998$$

- La variable X_2, con coeficiente -.924 recibe el mayor peso en la componente Y_1. también tiene la correlación más grande (en valor absoluto) con Y_1. La correlación de X_1 con Y_1, .925 es casi tan grande como la de X_2 con Y_1, indicando con eso que las dos variables son igualmente de importantes a la primera componente principal. Los tamaños relativos de los coeficientes de X_1 y X_2 sugieren asimismo, que X_2 contribuye más a la determinación de Y_1 que X_1.

Ejemplo: continuación

- Puesto que en este caso, ambos coeficientes son razonablemente grandes y tienen signos opuestos, se puede argumentar que ambas variables ayudan en la interpretación de Y_1.
- Finalmente observemos que:

$$\rho_{Y_2,X_1} = \rho_{Y_2,X_2} = 0 \quad \text{and} \quad \rho_{Y_2,X_3} = \frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} = 1$$ (como se esperaba)

- Las restantes correlaciones se omiten, puesto que la tercera componente no es importante.

Trabajando con SPSS

Introducción

- La utilidad principal del análisis factorial, es tomar una gran cantidad de instancias observables para medir una construcción o construcciones no observables. Por ejemplo: **"Asiste a fiestas ruidosas"**, **"habla mucho"**, **"parece cómodo interactuando con casi cualquier persona"** y **"generalmente se ve con otros"** son cuatro comportamientos que se pueden observar que pueden medir una construcción inobservable llamada "saliente" (Cattell, 1994).

Uso del Análisis Factorial

- Se usa con mayor frecuencia para identificar un pequeño número de factores (p. Ej., Salientes) que se pueden usar para representar relaciones entre conjuntos de variables interrelacionadas (p. Ej., Los cuatro descriptores anteriores).
- Se requieren cuatro pasos básicos para realizar un análisis factorial:
- 1. Calcule una matriz de correlación de todas las variables que se utilizarán en el análisis.
- 2. Factores de extracción.
- 3. Rotar factores para crear una estructura de factores más comprensible.
- 4. Interpretar los resultados.

- El análisis factorial no comienza con una variable dependiente. Comienza con una medida de la cantidad total de variación observada (similar a la suma total de cuadrados) en todas las variables que se han designado para el análisis factorial. Tenga en cuenta que esta "variación" es un poco difícil de entender conceptualmente (¿De dónde flota toda esta variación? ¿Cómo se ve que la huele? ¿La capto?)

Procedimiento

- El primer paso en el análisis factorial es que la computadora seleccione la combinación de variables cuyas correlaciones compartidas explican la mayor cantidad de la varianza total. Esto se llama Factor 1 (o Componente 1: las palabras se usan indistintamente).

Segundo paso, tercer paso, etc.

- El análisis factorial extraerá un segundo factor. Esta es la combinación de variables que explica la mayor cantidad de varianza que queda, es decir, la variación después de que se extrae el primer factor. Esto se llama Factor 2 (o Componente 2). Este procedimiento continúa para un tercer factor, cuarto factor, quinto factor, y así sucesivamente, hasta que se hayan extraído tantos factores como variables.

- En el procedimiento SPSS predeterminado, a cada una de las variables se le asigna inicialmente un valor de comunalidad de 1.0. Las comunidades están diseñadas para mostrar la proporción de varianza que los factores contribuyen a explicar una variable particular. Estos valores varían de 0 a 1 y pueden interpretarse de manera similar a una R múltiple, donde 0 indica que los factores comunes no explican ninguna de la varianza en una variable particular, y 1 indica que toda la varianza en esa variable se explica por factores comunes. Sin embargo, para el procedimiento predeterminado en la fase de extracción inicial, a cada variable se le asigna una comunalidad de 1.0.

- Después de extraer el primer factor, SPSS imprime un valor propio a la derecha del número de factor (por ejemplo, Número de factor = 1; valor propio = 5.13). **Los valores propios están diseñados para mostrar la proporción de varianza explicada por cada factor** (no cada variable como lo hacen las comunalidades). **El primer valor propio siempre será mayor (y siempre será mayor que 1.0)** porque el primer factor (según la definición del procedimiento) siempre explica la mayor cantidad de varianza total. Luego enumera el porcentaje de la varianza explicada por este factor (el valor propio dividido por el número total de variables), y esto es seguido por un porcentaje acumulativo.

- Para cada factor sucesivo, el valor propio impreso será más pequeño que el anterior, y el porcentaje acumulado (**de la varianza explicada**) totalizará el 100% después de que se haya calculado el factor final. La ausencia de la palabra se demuestra significativamente por el hecho de que el comando Factor extrae tantos factores como variables, independientemente de si los factores posteriores explican o no una cantidad significativa de varianza adicional.

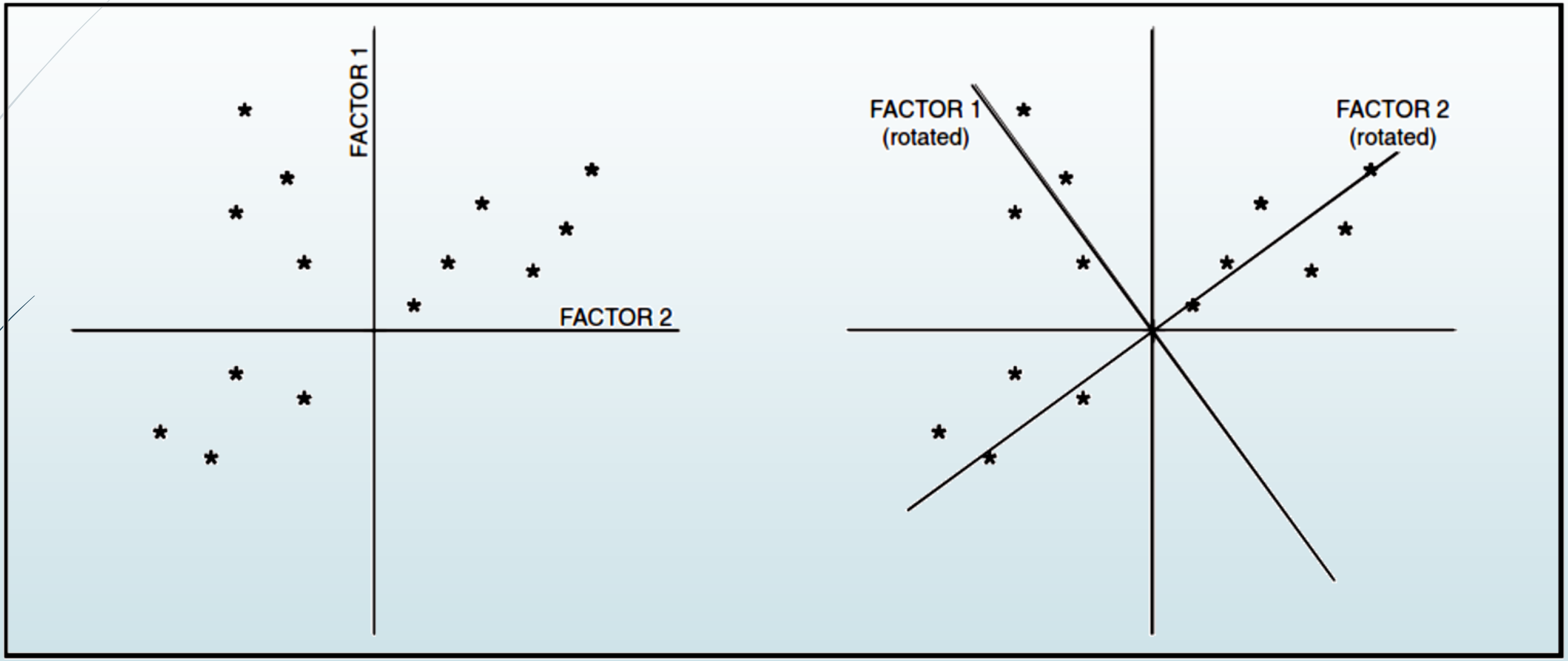
- Los factores extraídos por SPSS casi nunca son de interés para el investigador, si se tiene tantos factores como variables, pues no ha logrado entender para qué se creó el análisis factorial.
- El objetivo es explicar los fenómenos de interés con menos de la cantidad original de variables, generalmente son menos. ¿Recuerdas a Cattell? Comenzó con 4.500 descriptores y terminó con 16 rasgos.

- El primer paso es decidir qué factores desea conservar en el análisis. El criterio de sentido común para los factores de retención es que cada factor retenido debe tener algún tipo de validez aparente o validez teórica; pero antes del proceso de rotación, a menudo es imposible interpretar qué significa cada factor. Por lo tanto, el investigador generalmente selecciona un criterio matemático para determinar qué factores retener. El valor predeterminado de SPSS es mantener cualquier factor con un valor propio mayor que 1.0. Si un factor tiene un valor propio inferior a 1.0, explica menos varianza que una variable original y generalmente se rechaza.

- Recuerde, SPSS imprimirá tantos factores como variables y, por lo general, solo para unos pocos factores el valor propio será mayor que 1.0. Existen otros criterios para la selección (como el diagrama de pantalla) o razones conceptuales (según su conocimiento de los datos) que se pueden usar. El procedimiento para seleccionar un número diferente al predeterminado se describirá en la sección Paso a paso.

- **Una vez que los factores han sido seleccionados, el siguiente paso es rotarlos**. **La rotación es necesaria** porque la estructura factorial original es matemáticamente correcta pero es difícil de interpretar.
- El objetivo de la rotación es lograr lo que se llama estructura simple, es decir, cargas de factor alto en un factor y cargas bajas en todos los demás. Las cargas de factores varían entre ± 1.0 e indican la fuerza de la relación entre una variable particular y un factor particular, de manera similar a una correlación. Por ejemplo, la frase "disfruta de fiestas ruidosas" podría tener una carga alta en un factor "saliente" (quizás> .6) y una carga baja en un factor de "inteligencia" (quizás <.1).

- Esto se debe a que se cree que una declaración de fiestas divertidas está relacionada con la extrovertida, pero no está relacionada con la inteligencia. Idealmente, una estructura simple tendría una carga de variables completamente en un factor y nada en los demás. En el segundo gráfico (diapositiva siguiente), esto estaría representado por todos los asteriscos que están en las líneas de factores rotados.
- Sin embargo, en la investigación en ciencias sociales, esto nunca sucede, y el objetivo es rotar los ejes para que los puntos de datos estén lo más cerca posible de los ejes rotados. Los siguientes gráficos son buenas representaciones de cómo se vería una estructura no rotada y una estructura rotada. SPSS imprimirá gráficos de su estructura de factores (después de la rotación) e incluirá una tabla de coordenadas para mayor claridad.


FACTOR 1
FACTOR 2
FACTOR 1
(rotated)
FACTOR 2
(rotated)

- La rotación no altera la precisión matemática de la estructura factorial, al igual que mirar una imagen desde el frente en lugar de desde el lado no altera la imagen, y cambiar la medida de la altura en pulgadas a la altura en centímetros no altera la altura de la persona. Originalmente, la rotación se hacía a mano, y el investigador colocaría los ejes en la ubicación que parecía crear la estructura de factor óptima.
- La rotación manual no es posible con SPSS, pero hay varios procedimientos matemáticos disponibles para rotar los ejes a la mejor estructura simple. **Varimax** es el procedimiento predeterminado utilizado por SPSS, pero hay varios otros (mencionados en la sección Paso a paso).

- ROTACIONES OBLICUAS Las rotaciones Varimax se llaman rotaciones ortogonales porque los ejes que se giran permanecen en ángulo recto entre sí. A veces es posible lograr una mejor estructura simple al divergir de la perpendicular. Los procedimientos **Direct Oblimin** y **Promax** permiten al investigador desviarse de la ortogonal para lograr una mejor estructura simple. Conceptualmente, esta desviación significa que los factores ya no están sin correlación entre sí. Esto no es necesariamente perturbador, porque pocos factores en las ciencias sociales no están totalmente correlacionados. El uso de rotaciones oblicuas puede ser bastante complicado, y (aquí insertamos nuestro descargo de responsabilidad estándar) no debe intentar usarlas a menos que tenga una comprensión clara de lo que está haciendo.

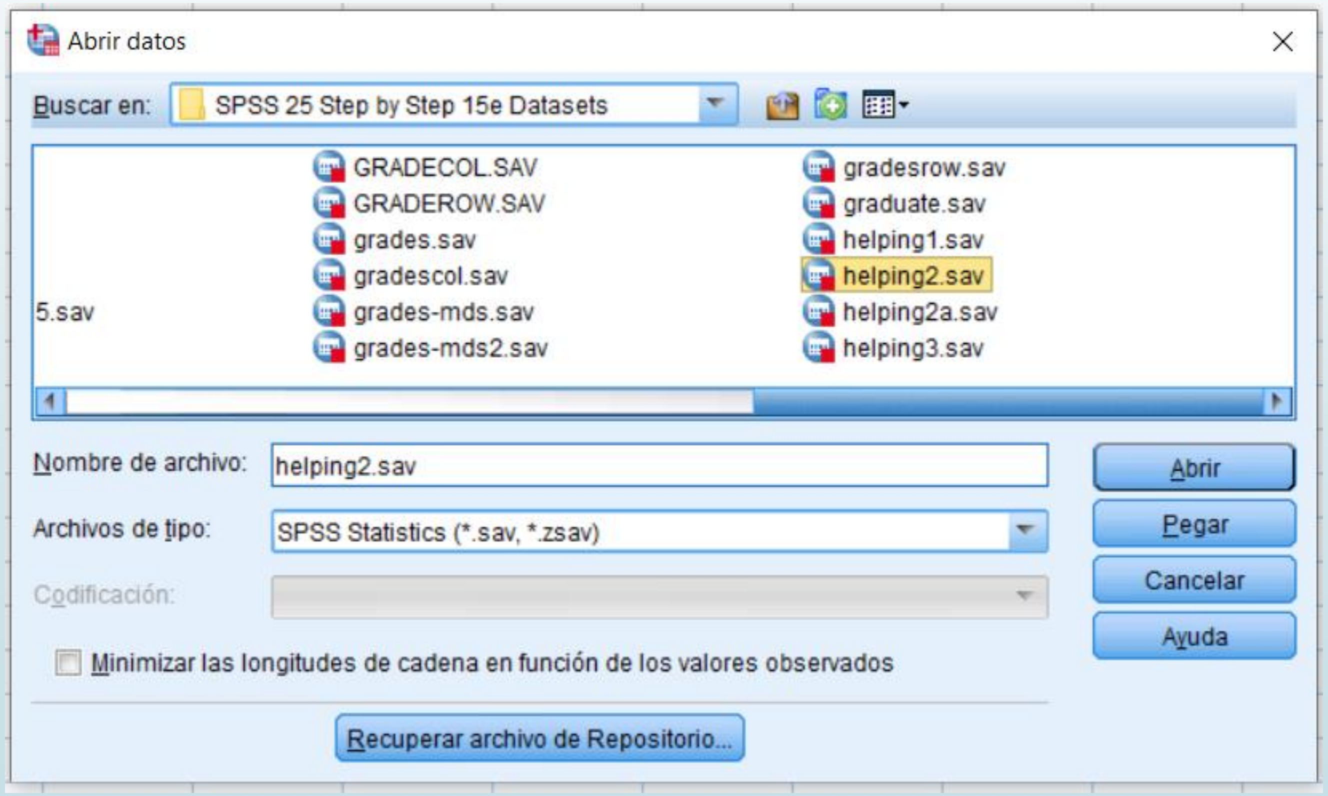
Vamos a ampliar:

- Probablemente no debería intentar llevar a cabo un análisis factorial a menos que haya tenido un curso y / o tenga una comprensión conceptual clara del procedimiento. La técnica para especificar la rotación directa de Oblimin o Promax se describe en la sección Paso a paso. No demostramos este procedimiento porque requiere más atención de la que podemos conceder aquí.

Paso por paso en Análisis de Factores

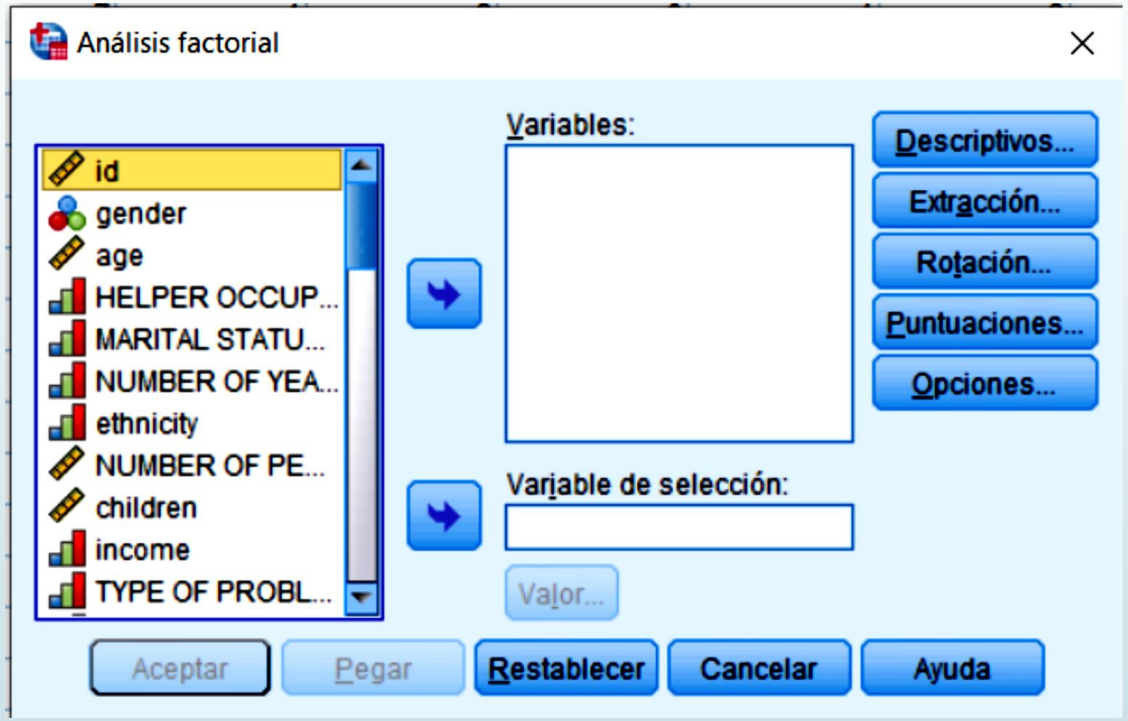
Iniciando

- Para iniciar, cargar el archivo helping2.sav siguiendo la secuencia:
- Archivo -> Abrir -> Datos -> helping2.sav -> Abrir



Iniciando

- Después desde el menú principal elegir:
- Analizar -> Reducción de dimensiones -> Factor
- Y aparece la siguiente ventana:



(Ventana 0)

- Lo que se muestra es una caja de variables a la izquierda, al centro una caja dinámica llamada **Variables**, y a la derecha 5 botones con diferentes opciones disponibles.

Iniciando

- Aunque la aritmética del análisis factorial es ciertamente compleja, se puede realizar un análisis factorial simplemente pegando algunas variables en el cuadro activo, seleccionando rotaciones Varimax y haciendo clic en Aceptar.

Iniciando

- De hecho, cualquier análisis factorial comenzará con el pegado de variables cuidadosamente seleccionadas en el cuadro Variables. Para el presente ejemplo, seleccionaremos las 15 preguntas de autoeficacia (efficacy 1, ..., efficacy 5) para su análisis. Las preguntas de eficacia se muestran en el cuadro de lista variable en la Ventana 0. Los cinco botones (a la derecha) representan muchas opciones disponibles y pueden procesarse en cualquier orden. Comenzamos con **Descriptivos**.

Descriptivos

- Al seleccionar Descriptivos, aparece la siguiente ventana:

- Describiremos solo aquellas selecciones que los investigadores usan con mayor frecuencia. La opción **Descriptivos univariantes** es bastante útil. Enumera en cuatro columnas ordenadas los nombres de las variables, las medias, las desviaciones estándar y las etiquetas de variables siempre útiles. Este cuadro será referido frecuentemente durante el curso del análisis.

Descriptivos

- La ***solución inicial*** se selecciona de forma predeterminada y enumera los nombres de las variables, las comunidades iniciales (1.0 de forma predeterminada), los factores, los valores propios y el porcentaje y el porcentaje acumulativo contabilizados por cada factor. La matriz de correlación es el punto de partida de cualquier análisis factorial.

Descriptivos

- Describimos cuatro de las estadísticas de uso frecuente relacionadas con la matriz de correlación:

• **Coeficientes**: esta es simplemente la matriz de correlación de las variables incluidas.

• **Niveles de significación**: estos son los valores de p asociados con cada correlación.

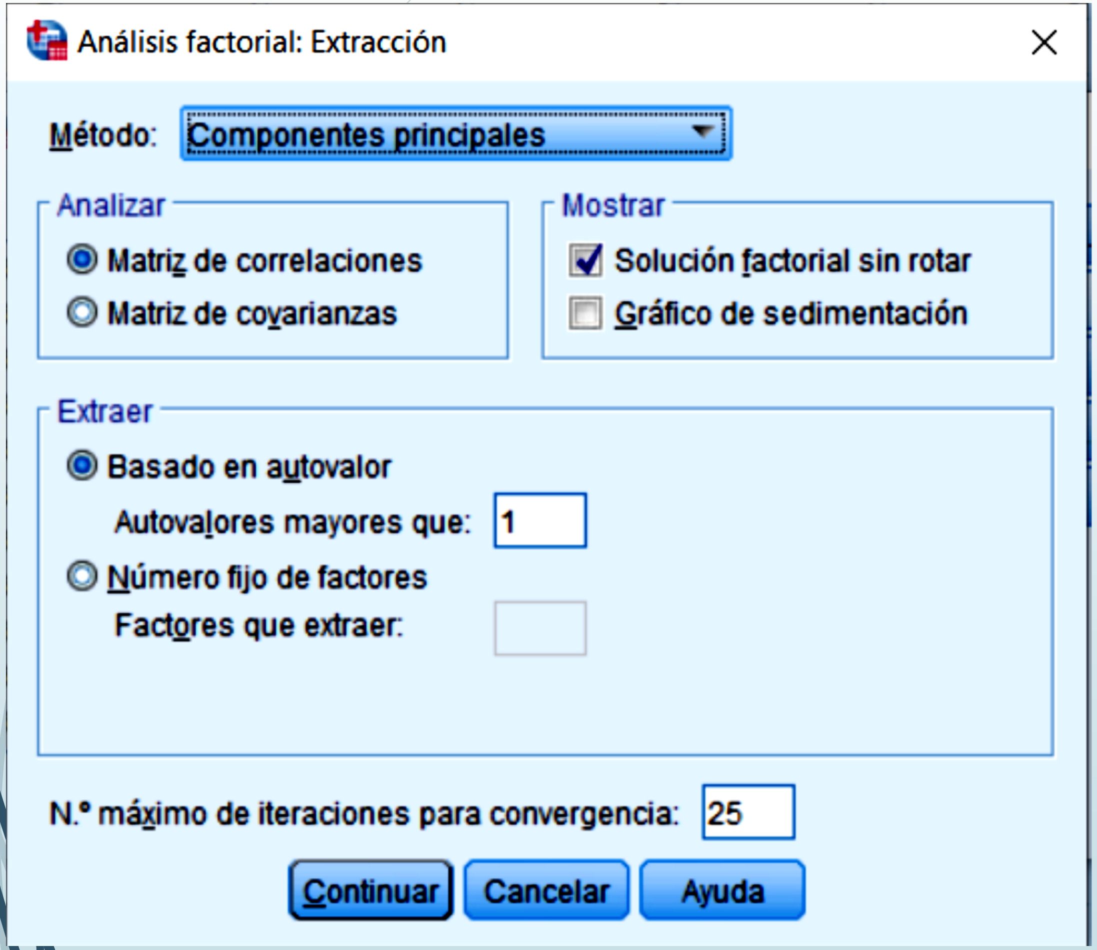
• **Determinante**: es el determinante de la matriz de correlación. Se utiliza en la computación de valores para pruebas de normalidad multivariante.

• **KMO y la prueba de esfericidad de Bartlett**: la prueba de KMO y la prueba de esfericidad de Bartlett son pruebas de normalidad multivariante y de adecuación de muestreo (la adecuación de sus variables para realizar el análisis factorial). Esta prueba está seleccionada por defecto.

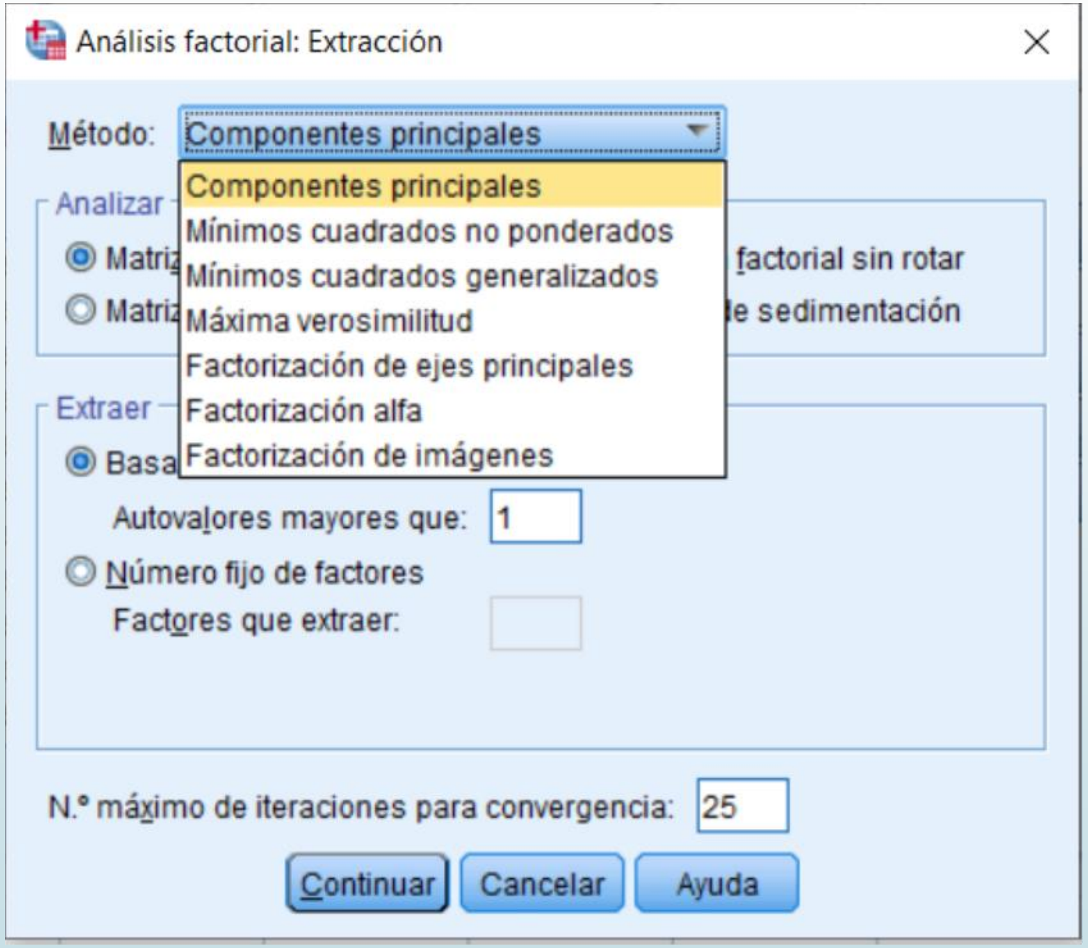
Extracción

- Un clic en el botón Extracción abre un nuevo cuadro de diálogo (Ventana 1) que trata sobre el método de extracción, los criterios para la selección de factores, la visualización de resultados relacionados con la extracción de factores y la especificación del número de iteraciones para el procedimiento para converger a una solución.

- El método de extracción de factores incluye siete opciones.



Ventana 1



Ventana 2

El método de extracción de factores incluye siete opciones.

- Los componentes principales son el procedimiento predeterminado; Un clic en el revela los otros seis (Ventana 2):
- • Mínimos cuadrados no ponderados
- • Mínimos cuadrados generalizados
- • Máxima verosimilitud
- • Factorización del eje principal
- • Factorización alfa
- • Factoring de imagen
- Sin embargo, el método de componentes principales es el más utilizado (con la máxima probabilidad el siguiente más común), y las limitaciones de espacio prohíben la discusión de las otras opciones.

Cuadro analizar

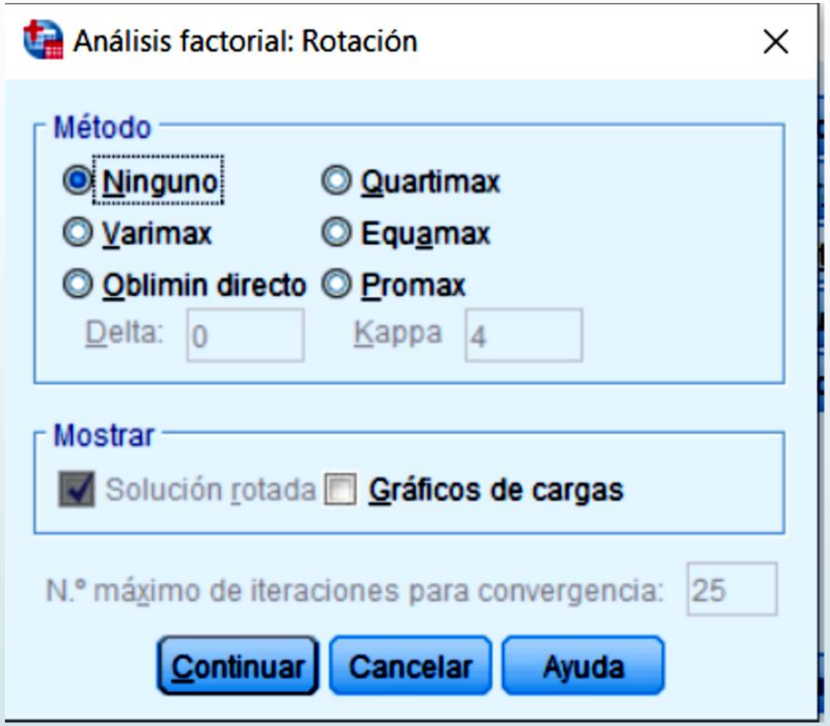
- El cuadro **Analizar** (Ver ventana 1) le permite seleccionar la **matriz de correlación** o la **matriz de covarianza** como punto de partida para su análisis. En **Extraer**, el **Número fijo de factores** para la rotación se basa en que el valor propio sea mayor que 1 (el valor predeterminado), la selección de un valor propio diferente o la simple identificación de cuántos factores desea seleccionar para la rotación.
- En la caja **Mostrar**, la **Solución factorial sin girar no girado** se selecciona de manera predeterminada, pero a menos que sea matemáticamente sofisticado, la solución de factorial sin girar rara vez revela mucho.
- La mayoría de los investigadores elegirían anular la selección de esta opción. También puede solicitar un **Diagrama de sedimentación** (ilustrado en la sección de Salida). Finalmente, puede designar el **número de iteraciones** que desea para la convergencia. El valor predeterminado de 25 es casi siempre suficiente.

Rotación

- Un clic en Rotación pasa al siguiente paso en el análisis factorial, rotando los factores a la solución final. La Ventana 3 muestra las opciones disponibles. Hay tres métodos ortogonales diferentes para la rotación, **Varimax** (el método más popular, sí, pero usarlo requiere resistir el desdén del factor analítico de élite), **Equama**x y **Quartimax**. Los procedimientos **Direct Oblimin** y **Promax** permiten una rotación no ortogonal de factores seleccionados. Los dos parámetros oblicuos de los procedimientos (δ y k) generalmente se pueden dejar en los valores predeterminados. Como se sugirió en la introducción, ni siquiera piense en intentar rotaciones oblicuas sin un curso sólido en el análisis factorial.

Rotación

- Para la visualización, la **Solución rotada** se selecciona de manera predeterminada y representa la esencia de lo que el análisis factorial está diseñado para hacer.
- Si desea una visualización de la estructura de factores después de la rotación, haga clic en la opción **Gráficos de carga** (ver Ventana 3). La opción Gráficos proporciona por defecto un gráfico tridimensional de los primeros tres factores (Ventana 4).


Análisis factorial: Rotación
Método
Ninguno
Quartimax
Varimax
Equamax
Oblimin directo
Promax
Delta:
0
Kappa
4
Mostrar
Solución rotada
Gráficos de cargas
N.º máximo de iteraciones para convergencia:
25
Continuar
Cancelar
Ayuda

Ventana 3

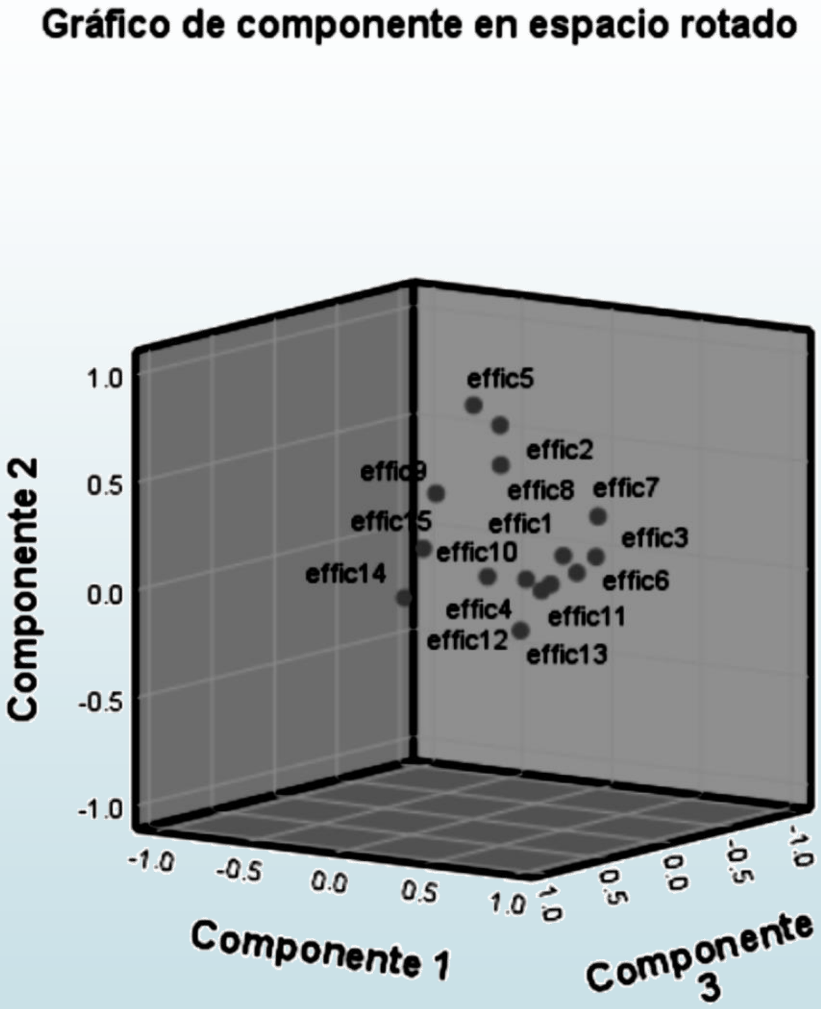
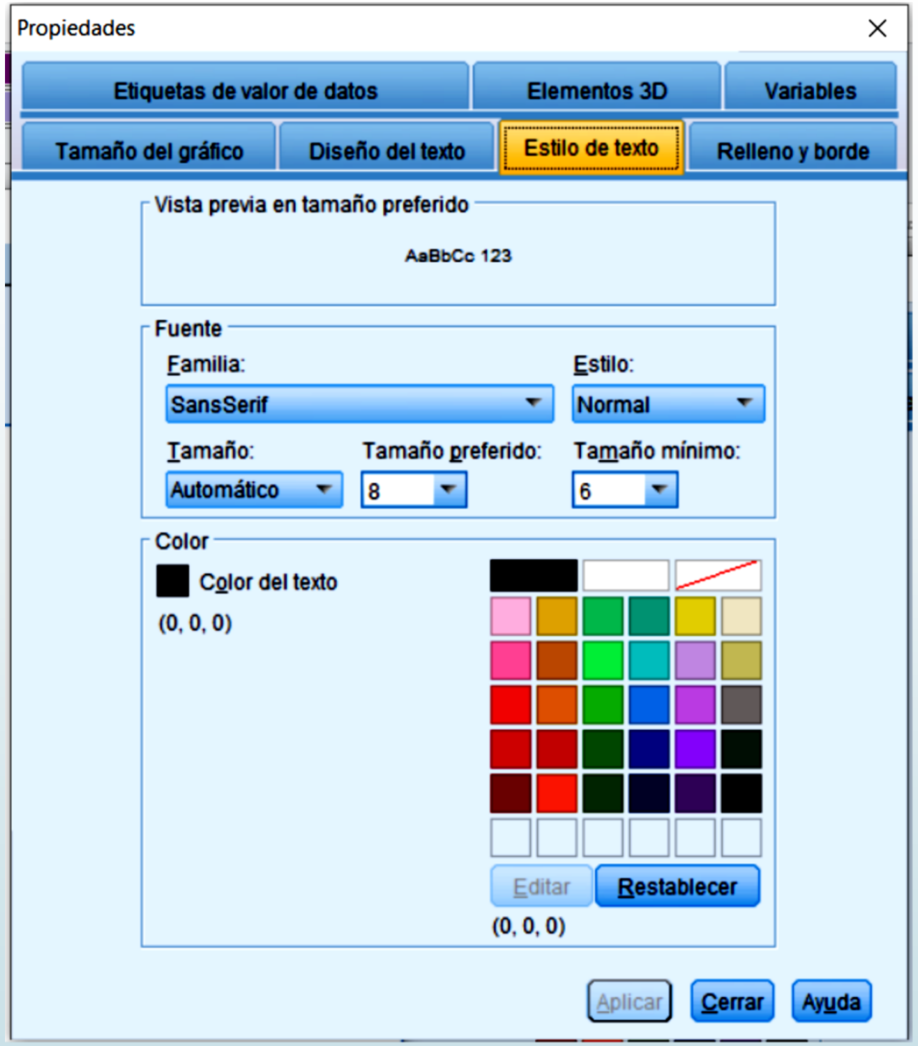


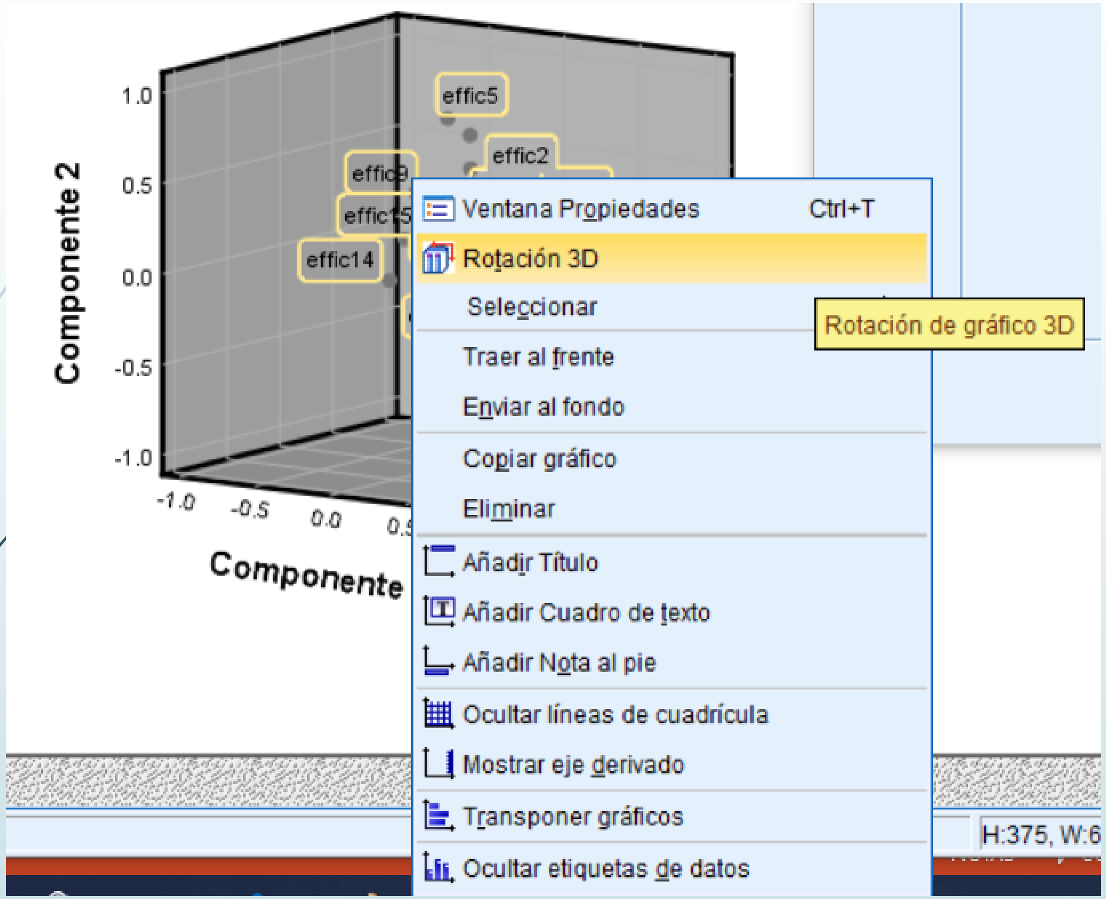
Gráfico de componente en espacio rotado
effic5
effic2
effic9
effic8
effic7
effic15
effic1
effic3
effic10
effic14
effic6
effic4
effic11
effic12
effic13
1.0
0.5
0.0
-0.5
-1.0
Componente 2
-1.0
-0.5
0.0
0.5
1.0
Componente 1
1.0
0.5
0.0
-0.5
-1.0
Componente 3

Ventana 4

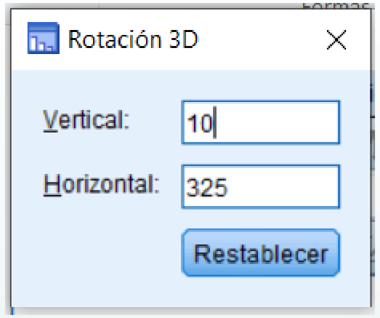
- Los gráficos bidimensionales son mucho más fáciles de interpretar. Las opciones de Análisis factorial de SPSS no incluyen esta posibilidad. Sin embargo, si desea convertir el gráfico 3-D en un gráfico 2-D mucho más simple, simplemente abra el gráfico haciendo doble clic en él. Una vez que aparece el gráfico sujeto a edición (Ventana 5), haga clic con el botón derecho en el gráfico y se abrirá un cuadro de diálogo llamado Rotación 3-D (Ventanas 6a y 6b). Luego gire el tercer eje (Z) fuera de la imagen. Mientras edita para mayor claridad, tenga paciencia mientras "lucha con la bestia". Estos gráficos en 2-D son bastante interpretables, especialmente si hace referencia a los valores enumerados en la matriz de componentes (o factores) rotados.


Propiedades
Etiquetas de valor de datos
Elementos 3D
Variables
Tamaño del gráfico
Diseño del texto
Estilo de texto
Relleno y borde
Vista previa en tamaño preferido
AaBbCc 123
Fuente
Familia:
SansSerif
Estilo:
Normal
Tamaño:
Automático
Tamaño preferido:
8
Tamaño mínimo:
6
Color
Color del texto
(0, 0, 0)
Editar
Restablecer
(0, 0, 0)
Aplicar
Cerrar
Ayuda

Ventana 5

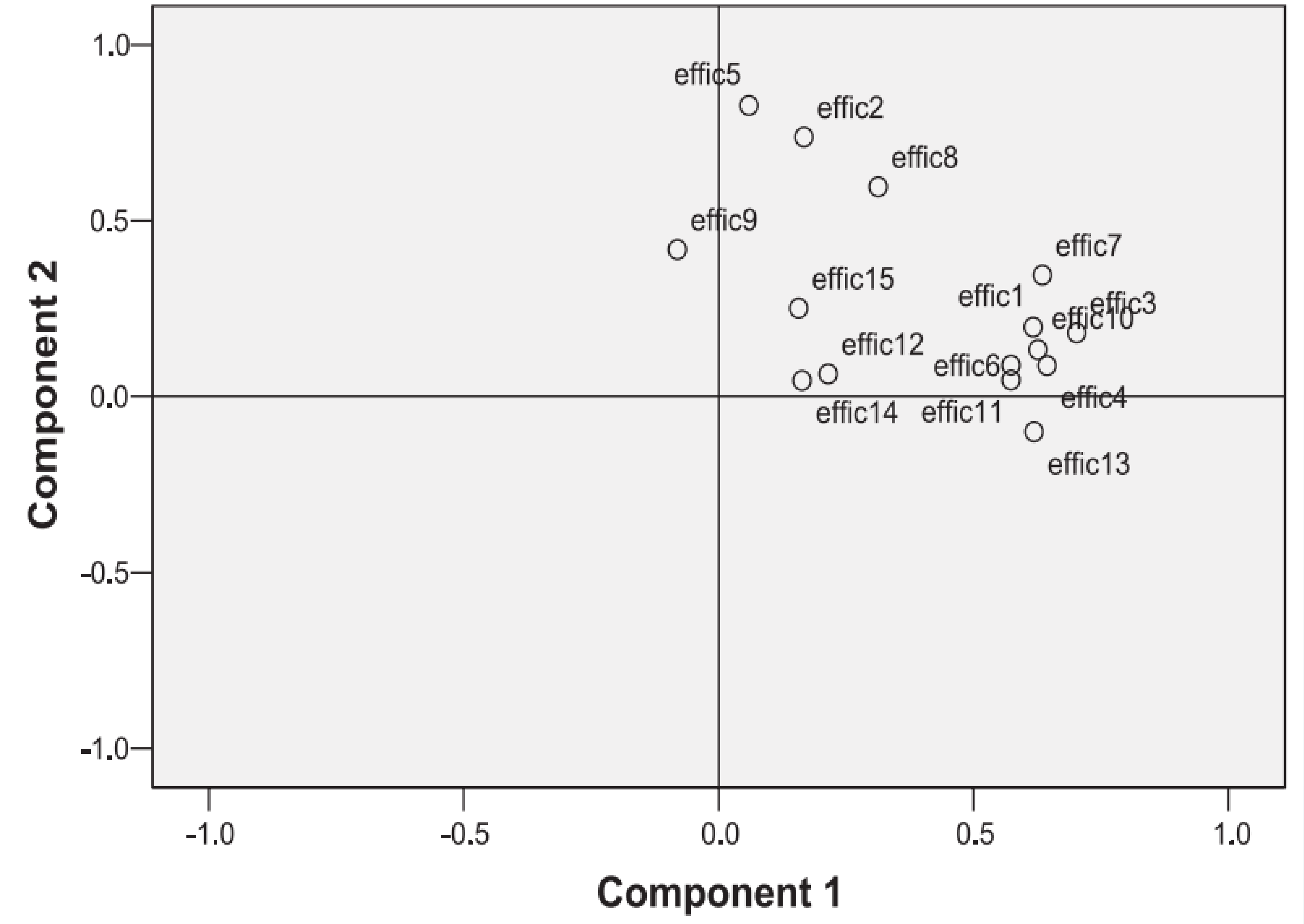


Ventana 6a



Ventana 6b

- Observe que en el cuadro (Ventana 7), el eje horizontal es el Factor 1 (eficacia para los tipos de ayuda empáticos) y el eje vertical es el Factor 2 (eficacia para los tipos de ayuda informativos). Tenga en cuenta también que los ocho puntos de datos en el extremo derecho presentan altas cargas (.6 a .8) sobre la eficacia para la ayuda empática (Factor 1) y cargas bajas (−.1 a .4) sobre la eficacia para la ayuda informativa (Factor 2).

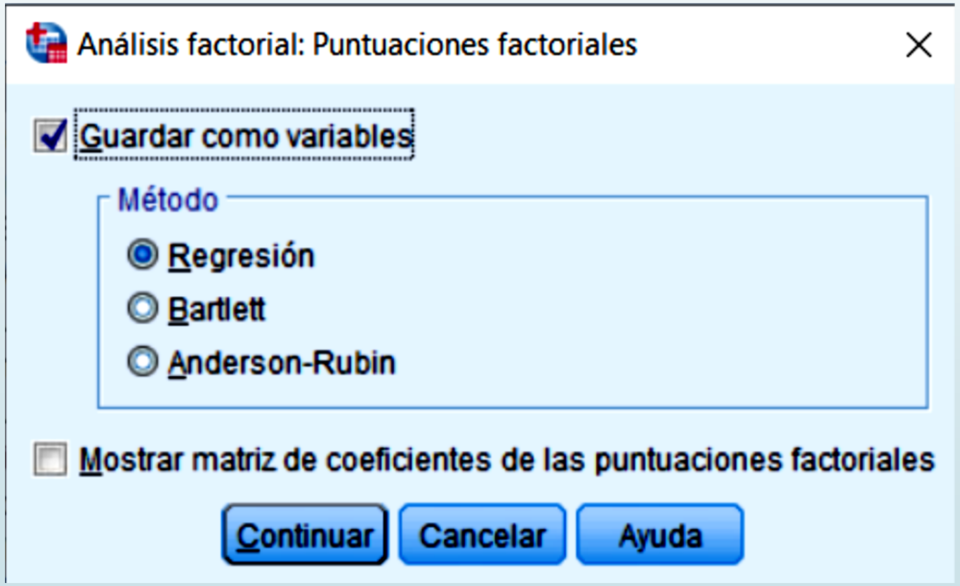


effic5
effic2
effic8
effic9
effic7
effic15
effic1
effic3
effic10
effic12
effic6
effic4
effic14
effic11
effic13
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
Component 2
Component 1

Ventana 7

Puntuaciones

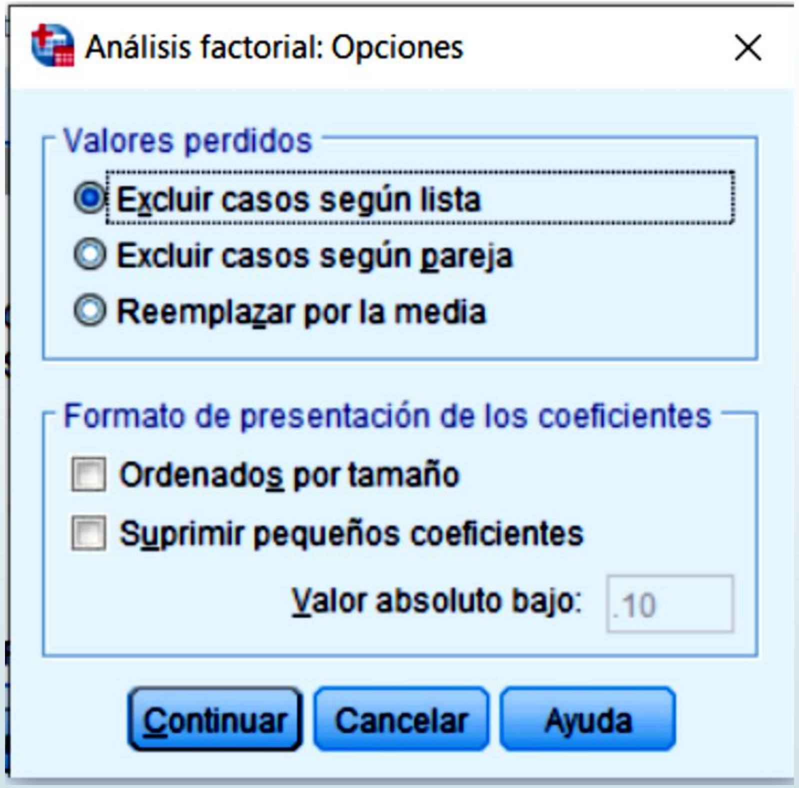
- Volviendo a la Ventana 0: un clic en el botón **Puntuaciones** abre un pequeño cuadro de diálogo (Ventana 8) que le permite guardar ciertas puntuaciones como variables. Esta ventana no se muestra, pero la matriz del coeficiente de puntuación del factor de visualización (cuando se selecciona) muestra la matriz del coeficiente de puntuación del componente en la salida.



Ventana 8

Opciones

- Finalmente, un clic en el botón **Opciones** (de Ventana 0) abre un cuadro de diálogo (Ventana 9) que permite dos opciones diferentes relacionadas con la visualización de la matriz de factores rotados. La selección de Ordenar por tamaño es muy útil. Ordena las variables por la magnitud de sus cargas factoriales por factor. Por lo tanto, si seis variables se cargan en el Factor 1, las cargas de factor para esas seis variables se enumerarán de la más grande a la más pequeña en la columna titulada Factor 1. Lo mismo será cierto para las variables que se cargan altamente en el segundo factor, el tercer factor , Etcétera.


Análisis factorial: Opciones
Valores perdidos
Excluir casos según lista
Excluir casos según pareja
Reemplazar por la media
Formato de presentación de los coeficientes
Ordenados por tamaño
Suprimir pequeños coeficientes
Valor absoluto bajo:
.10
Continuar
Cancelar
Ayuda

Ventana 9

- La sección Salida muestra esta característica. Entonces puede suprimir cargas de factores que son menores que un valor particular (el valor predeterminado es .10) si considera que tales cargas son insignificantes. En realidad, puedes hacerlo sin importar cómo te sientas. Cualquiera que sea capaz de realizar un análisis factorial ya se ha ocupado de los valores perdidos y se sentiría insultado por la implicación de que él o ella recurriría a un procedimiento automático en esta etapa del proceso.

Análisis factorial a las 15 variables (efficacy) con el método Varimax

El análisis factorial realiza lo siguiente:

1. Calcula una matriz de correlación de las 15 preguntas de eficacia del archivo de datos,

2. Extrae 15 factores por el método de componentes principales,

3. Selecciona como factores a rotar todos los factores que tienen un valor propio mayor que 1.0,

4. Rota los factores seleccionados a una solución Varimax, y

5. Imprime la matriz de transformación de factores.

Análisis factorial

Varianza total explicada

Componente	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	5.133	34.221	34.221	5.133	34.221	34.221	3.358	22.388	22.388
2	1.682	11.211	45.432	1.682	11.211	45.432	2.065	13.766	36.154
3	1.055	7.030	52.462	1.055	7.030	52.462	1.803	12.023	48.177
4	1.028	6.851	59.313	1.028	6.851	59.313	1.670	11.136	59.313
5	.885	5.902	65.215						
6	.759	5.057	70.272						
7	.628	4.185	74.457						
8	.624	4.157	78.614						
9	.589	3.927	82.541						
10	.530	3.530	86.072						
11	.494	3.294	89.366						
12	.468	3.120	92.486						
13	.429	2.863	95.349						
14	.398	2.651	98.000						
15	.300	2.000	100.000						

Método de extracción: análisis de componentes principales.

Prueba de KMO y Bartlett[a]

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		.871
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	1321.696
	gl	105
	Sig.	.000

a. Se basa en correlaciones

Varianza total explicada

	Componente	Autovalores iniciales[a]			Sumas de cargas al cuadrado de la extracción			Sumas de cargas al cuadrado de la rotación		
		Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
Puro	1	15.107	32.731	32.731	15.107	32.731	32.731	9.372	20.305	20.305
	2	5.748	12.453	45.183	5.748	12.453	45.183	7.473	16.191	36.496
	3	3.851	8.344	53.527	3.851	8.344	53.527	6.035	13.076	49.572
	4	3.683	7.980	61.507	3.683	7.980	61.507	5.509	11.935	61.507
	5	2.799	6.063	67.570						
	6	2.503	5.424	72.994						
	7	2.249	4.873	77.866						
	8	1.790	3.878	81.744						
	9	1.690	3.661	85.405						
	10	1.506	3.262	88.667						
	11	1.372	2.973	91.640						
	12	1.219	2.641	94.282						
	13	1.030	2.232	96.513						
	14	.901	1.953	98.466						
	15	.708	1.534	100.000						
Reescalado	1	15.107	32.731	32.731	4.977	33.178	33.178	3.708	24.722	24.722
	2	5.748	12.453	45.183	1.684	11.228	44.407	2.032	13.547	38.269
	3	3.851	8.344	53.527	1.001	6.672	51.079	1.668	11.121	49.391
	4	3.683	7.980	61.507	.977	6.517	57.596	1.231	8.205	57.596
	5	2.799	6.063	67.570						
	6	2.503	5.424	72.994						
	7	2.249	4.873	77.866						
	8	1.790	3.878	81.744						
	9	1.690	3.661	85.405						
	10	1.506	3.262	88.667						
	11	1.372	2.973	91.640						
	12	1.219	2.641	94.282						
	13	1.030	2.232	96.513						
	14	.901	1.953	98.466						
	15	.708	1.534	100.000						

Método de extracción: análisis de componentes principales.

a. Al analizar una matriz de covarianzas, los autovalores iniciales son los mismos entre la solución re-escalada y pura.

Matriz de componentea

	Componente 1	2	3	4
EFFICACY FOR ENCOURAGE REASSURE	.704	-.051	.074	-.173
EFFICACY FOR TASKS OR SERVICES	.514	.561	-.220	-.026
EFFICACY FOR APPRAISE/CLARIFY	.651	-.119	-.048	-.332
EFFICACY FOR VALIDATE AFFIRM	.618	-.327	-.151	-.060
EFFICACY FOR LOANING MATEIALS	.458	.613	-.343	.089
EFFICACY FOR INFORMATION ADVICE	.665	.069	.392	-.319
EFFICACY FOR EXPRESS WILLINGNESS TO HELP	.640	.087	-.087	-.385
EFFICACY FOR PARTICIPATE IN ACTIVITIES	.545	.210	-.400	.077
EFFICACY FOR FIND SOMEONE TO HELP	.407	.556	.314	.196
EFFICACY FOR EXPRESS SYMPATHY EMPATHY CONCERN	.645	-.396	-.290	.105
EFFICACY FOR REDUCE TENSION TELL JOKES	.598	-.246	.019	-.057
EFFICACY FOR TEACH TO DO BETTER	.523	.206	.561	.028
EFFICACY FOR EMPATHIC LISTENING	.650	-.465	.054	.112
EFFICACY FOR RELIEVE OF SELF BLAME	.553	-.162	.182	.579
EFFICACY FOR OPEN-ENDED QUESTION	.517	-.021	-.048	.468

Método de extracción: análisis de componentes principales.
a 4 componentes extraídos.

Matriz de transformación de componente

Componente	1	2	3	4
1	.733	.405	.418	.352
2	-.441	.736	-.295	.420
3	-.113	-.538	.017	.835
4	-.506	.066	.859	-.044

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.

Comunalidades

	Puro		Reescalado	
	Inicial	Extracción	Inicial	Extracción
EFFICACY FOR ENCOURAGE REASSURE	2.311	1.133	1.000	.490
EFFICACY FOR TASKS OR SERVICES	3.204	1.796	1.000	.561
EFFICACY FOR APPRAISE/CLARIFY	1.870	.783	1.000	.419
EFFICACY FOR VALIDATE AFFIRM	2.467	1.226	1.000	.497
EFFICACY FOR LOANING MATEIALS	4.206	3.218	1.000	.765
EFFICACY FOR INFORMATION ADVICE	2.192	1.059	1.000	.483
EFFICACY FOR EXPRESS WILLINGNESS TO HELP	2.315	1.042	1.000	.450
EFFICACY FOR PARTICIPATE IN ACTIVITIES	3.936	2.228	1.000	.566
EFFICACY FOR FIND SOMEONE TO HELP	4.582	3.836	1.000	.837
EFFICACY FOR EXPRESS SYMPATHY EMPATHY CONCERN	2.336	1.310	1.000	.561
EFFICACY FOR REDUCE TENSION TELL JOKES	2.536	1.157	1.000	.456
EFFICACY FOR TEACH TO DO BETTER	3.068	1.497	1.000	.488
EFFICACY FOR EMPATHIC LISTENING	2.404	1.449	1.000	.603
EFFICACY FOR RELIEVE OF SELF BLAME	3.704	1.965	1.000	.530
EFFICACY FOR OPEN-ENDED QUESTION	5.026	4.692	1.000	.934

Método de extracción: análisis de componentes principales.

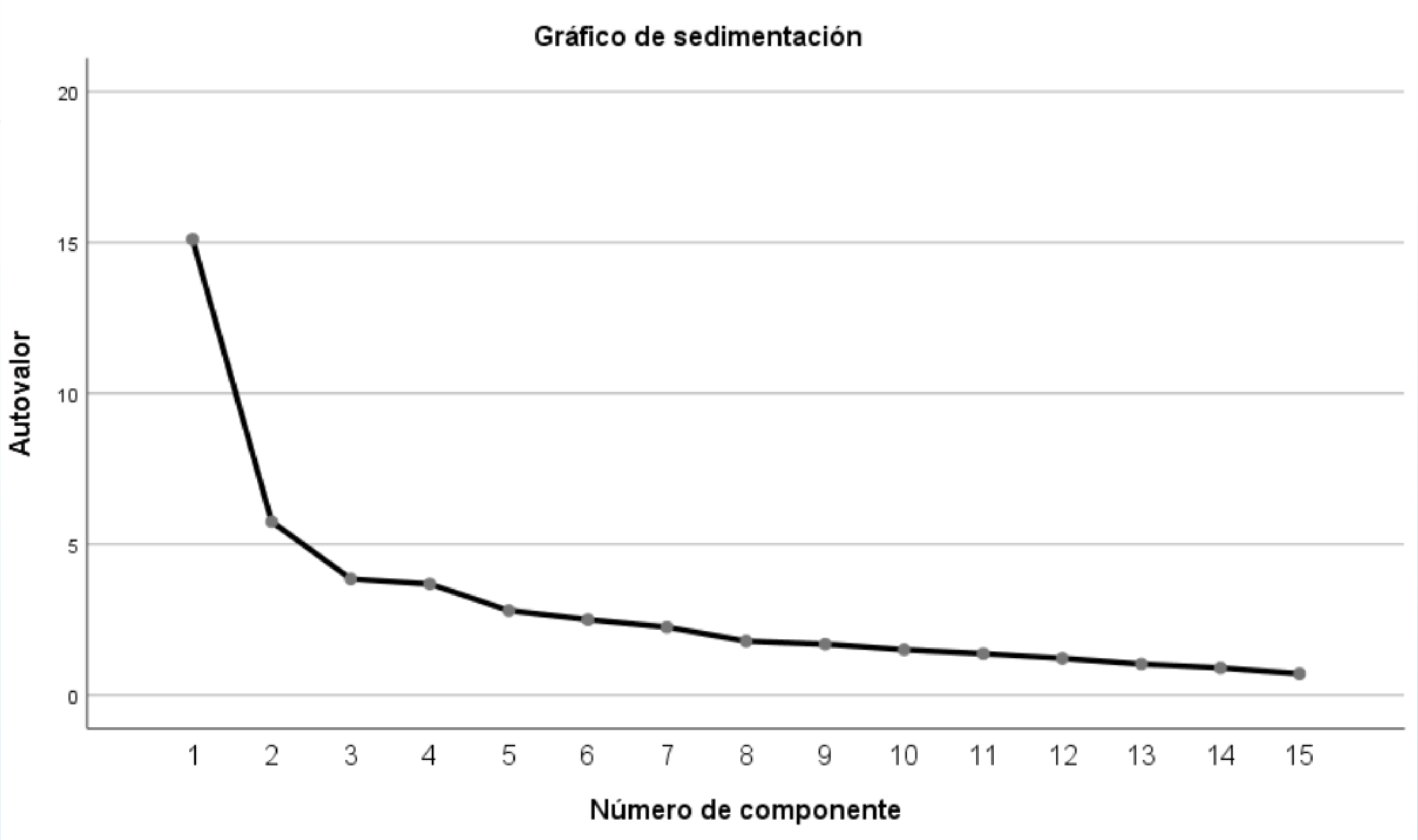


Gráfico de sedimentación
Autovalor
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Número de componente

Matriz de componente[a]

	Puro				Reescalado			
	Componente				Componente			
	1	2	3	4	1	2	3	4
EFFICACY FOR ENCOURAGE REASSURE	.991	.209	.294	.145	.652	.137	.194	.095
EFFICACY FOR INFORMATION ADVICE	.911	.077	.169	.441	.615	.052	.114	.298
EFFICACY FOR EXPRESS WILLINGNESS TO HELP	.909	.022	.463	-.037	.597	.014	.304	-.024
EFFICACY FOR PARTICIPATE IN ACTIVITIES	1.175	-.295	.594	-.638	.592	-.149	.300	-.321
EFFICACY FOR EXPRESS SYMPATHY EMPATHY CONCERN	.897	.641	.210	-.223	.587	.420	.137	-.146
EFFICACY FOR RELIEVE OF SELF BLAME	1.116	.534	-.534	.385	.580	.278	-.277	.200
EFFICACY FOR EMPATHIC LISTENING	.894	.735	.206	.259	.576	.474	.133	.167
EFFICACY FOR APPRAISE/CLARIFY	.787	.244	.323	.008	.575	.178	.236	.006
EFFICACY FOR TASKS OR SERVICES	1.020	-.798	.083	-.334	.570	-.446	.046	-.187
EFFICACY FOR VALIDATE AFFIRM	.868	.563	.395	.004	.553	.358	.252	.003
EFFICACY FOR REDUCE TENSION TELL JOKES	.873	.509	.361	.075	.548	.320	.227	.047
EFFICACY FOR TEACH TO DO BETTER	.938	-.137	-.035	.773	.536	-.078	-.020	.441
EFFICACY FOR LOANING MATEIALS	1.128	-1.179	.123	-.735	.550	-.575	.060	-.358
EFFICACY FOR FIND SOMEONE TO HELP	1.074	-1.202	-.453	1.016	.502	-.562	-.212	.474
EFFICACY FOR OPEN-ENDED QUESTION	1.326	.497	-1.476	-.713	.591	.222	-.658	-.318

Método de extracción: análisis de componentes principales.

a. 4 componentes extraídos.

Matriz de componente rotado[a]

	Puro				Reescalado			
	Componente				Componente			
	1	2	3	4	1	2	3	4
EFFICACY FOR EMPATHIC LISTENING	1.133	-.128	.258	.285	.731	-.083	.167	.184
EFFICACY FOR VALIDATE AFFIRM	1.087	.143	.070	.143	.692	.091	.045	.091
EFFICACY FOR EXPRESS SYMPATHY EMPATHY CONCERN	1.053	.192	-.088	.396	.689	.126	-.057	.259
EFFICACY FOR REDUCE TENSION TELL JOKES	1.046	.136	.157	.138	.657	.085	.099	.087
EFFICACY FOR ENCOURAGE REASSURE	.920	.342	.391	.134	.605	.225	.257	.088
EFFICACY FOR APPRAISE/CLARIFY	.811	.294	.176	.084	.593	.215	.129	.061
EFFICACY FOR EXPRESS WILLINGNESS TO HELP	.822	.551	.248	-.037	.540	.362	.163	-.024
EFFICACY FOR INFORMATION ADVICE	.749	.206	.669	.088	.506	.139	.452	.059
EFFICACY FOR LOANING MATEIALS	.056	1.753	.319	.198	.027	.855	.156	.097
EFFICACY FOR TASKS OR SERVICES	.215	1.229	.454	.182	.120	.687	.254	.102
EFFICACY FOR PARTICIPATE IN ACTIVITIES	.834	1.237	-.023	.036	.420	.624	-.012	.018
EFFICACY FOR FIND SOMEONE TO HELP	-.155	.699	1.817	.149	-.072	.327	.849	.070
EFFICACY FOR TEACH TO DO BETTER	.566	.144	1.069	.113	.323	.082	.610	.065
EFFICACY FOR OPEN-ENDED QUESTION	.439	.444	.120	2.070	.196	.198	.054	.924
EFFICACY FOR RELIEVE OF SELF BLAME	.826	-.092	.687	.895	.429	-.048	.357	.465

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.

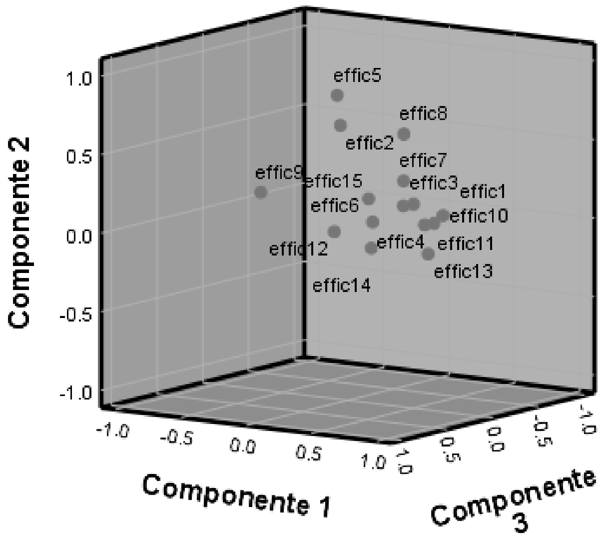
a. La rotación ha convergido en 6 iteraciones.

Matriz de transformación de componente

Componente	1	2	3	4
1	.657	.504	.423	.368
2	.592	-.653	-.383	.277
3	.463	.189	-.214	-.839
4	.061	-.533	.793	-.289

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.

Gráfico de componente en espacio rotado



Matriz de coeficiente de puntuación de componente[a]

	Componente 1	2	3	4
EFFICACY FOR ENCOURAGE REASSURE	.156	.004	.043	-.063
EFFICACY FOR TASKS OR SERVICES	-.060	.317	.009	-.010
EFFICACY FOR APPRAISE/CLARIFY	.134	.018	-.014	-.055
EFFICACY FOR VALIDATE AFFIRM	.225	-.025	-.054	-.060
EFFICACY FOR LOANING MATEIALS	-.143	.582	-.113	.003
EFFICACY FOR INFORMATION ADVICE	.111	-.050	.157	-.067
EFFICACY FOR EXPRESS WILLINGNESS TO HELP	.147	.085	-.015	-.114
EFFICACY FOR PARTICIPATE IN ACTIVITIES	.162	.385	-.234	-.129
EFFICACY FOR FIND SOMEONE TO HELP	-.246	.007	.758	-.027
EFFICACY FOR EXPRESS SYMPATHY EMPATHY CONCERN	.194	-.001	-.118	.037
EFFICACY FOR REDUCE TENSION TELL JOKES	.215	-.035	-.021	-.062
EFFICACY FOR TEACH TO DO BETTER	.062	-.117	.357	-.064
EFFICACY FOR EMPATHIC LISTENING	.223	-.126	.032	-.012
EFFICACY FOR RELIEVE OF SELF BLAME	.088	-.203	.208	.268
EFFICACY FOR OPEN-ENDED QUESTION	-.180	.041	-.151	.972

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.
Puntuaciones de componente.

a. Los coeficientes se han estandarizado.

Matriz de covarianzas de puntuación de componente

Componente	1	2	3	4
1	1.000	.000	.000	.000
2	.000	1.000	.000	.000
3	.000	.000	1.000	.000
4	.000	.000	.000	1.000

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.
Puntuaciones de componente.

- La reacción inicial del investigador que realizó este análisis fue "¡una estructura factorial bastante buena!" El primer factor se compone principalmente de variables que miden la eficacia para los tipos emocionales de ayuda. Una pregunta de ayuda instrumental ("¿Hasta qué punto tenía la capacidad de escuchar con atención o de evaluar la situación de su amigo?") Y otra de ayuda informativa ("¿Creía que era capaz de reducir la tensión y ayudar a su amigo a obtener su / su mente fuera del problema? ”) fueron incluidos en el primer factor. No es difícil ver por qué estos dos elementos podrían cargarse en el mismo factor que la eficacia para la ayuda emocional.

- La pregunta abierta (eficiencia 15) no se carga a ninguno de los cuatro factores. El factor 2 se compone completamente de las tres medidas restantes de eficacia para los tipos de donaciones informativas. Factor 3 se compone completamente de las tres medidas restantes de eficacia para los tipos de ayuda instrumental. El factor 4 es un factor bastante extraño y probablemente no se usaría. Se incluye la eficiencia14, una medida algo extraña con la que varios sujetos parecían confundidos. Se ocupó de la eficacia para ayudar al amigo a reducir la autoculpa.

- En muchas situaciones problemáticas, la culpa no era un problema. Este es el tipo de pensamiento que hace un investigador cuando intenta interpretar los resultados de un análisis factorial. El resultado actual parece producir un patrón bastante interpretable de tres tipos de eficacia: eficacia para los tipos emocionales de ayuda, eficacia para los tipos instrumentales de ayuda y eficacia para los tipos informativos de ayuda. El factor 4, el extraño, probablemente se descartaría.

¡Es Todo!

Ufff!!!!!!!!!!