



# COMPONENTES PRINCIPALES y ANALISIS FACTORIAL

**M<sup>a</sup> Purificación VICENTE GALINDO**

**[purivg@usal.es](mailto:purivg@usal.es)**

**M<sup>a</sup> Purificación GALINDO VILLARDÓN**

**[pgalindo@usal.es](mailto:pgalindo@usal.es)**

**Departamento de Estadística.**

**Universidad de Salamanca. España**

UN EJEMPLO CLÁSICO

## EJEMPLO

Los datos siguientes corresponden a 8 alumnos en las asignaturas de Matemáticas, Ciencias Naturales, Francés y Latín.

Alumno	Matem.	C. Nat.	Francés	Latín
1	9	8	6	7
2	10	9	10	10
3	3	5	9	8
4	9	9	8	8
5	7	6	3	5
6	5	5	5	5
7	5	5	7	6
8	4	4	3	4

Es bien conocida la división clásica de las asignaturas de la enseñanza media en asignaturas de Ciencias y asignaturas de Letras. En líneas generales, las primeras se caracterizan por un factor racional y empírico, mientras que las segundas tienen un significado más especulativo, siendo la memoria una de sus características más importantes. Cada una de las asignaturas tendrá un poco de ambas componentes, aunque será mayoritariamente de uno de los dos grupos. Por ejemplo, las Ciencias Naturales son consideradas de Ciencias pero es indudable que también tienen una importante componente de memoria.

Además, cada asignatura tendrá también un componente propio, más allá de lo que tengan de Ciencias o Letras.

## ¿Será adecuado pensar en ajustar un modelo Factorial?

*Para ello las variables observables tienen que estar correlacionadas*

The image shows the 'Análisis factorial' (Factorial Analysis) dialog box in SPSS, with the 'Descriptivos...' (Descriptives...) sub-dialog open. The 'Variables:' list contains 'Matematicas [Matematicas]', 'C Naturales [Naturales]', 'Francés [Francés]', and 'Latín [Latín]'. The 'Variable de selección:' (Selection variable) is empty. The 'Descriptivos...' sub-dialog has a title bar 'Descriptivos...' and three buttons: 'Rotación...', 'Puntuaciones...', and 'Opciones...'. The 'Descriptivos univariados' (Univariate Descriptives) section is checked, and 'Solución inicial' (Initial solution) is checked with the text 'Sin rotar' (No rotation) in red. The 'Matriz de correlaciones' (Correlation Matrix) section is also checked, and the following options are checked: 'Coeficientes' (Coefficients), 'Niveles de significación' (Significance levels), 'Determinante' (Determinant), 'KMO y prueba de esfericidad de Bartlett' (KMO and Bartlett's sphericity test), 'Inverso' (Inverse), 'Reproducida' (Reproduced), and 'Anti-imagen' (Anti-image).

Análisis factorial

Variables:

- Matematicas [Matematicas]
- C Naturales [Naturales]
- Francés [Francés]
- Latín [Latín]

Variable de selección:

Valor...

Rotación...

Puntuaciones...

Opciones...

**Descriptivos...**

☒ Descriptivos univariados

☒ Solución inicial **Sin rotar**

Matriz de correlaciones

- ☒ Coeficientes
- ☒ Niveles de significación
- ☒ Determinante
- ☒ KMO y prueba de esfericidad de Bartlett
- ☐ Inverso
- ☒ Reproducida
- ☒ Anti-imagen

**Descriptivos univariados** univariantes

☒ Descriptivos univariados

☒ Solución inicial

**Matriz de correlaciones**

☒ Coeficientes ☐ Inverso

☒ Niveles de significación ☒ Reproducida

☒ Determinante ☒ Anti-imagen

☒ KMO y prueba de esfericidad de Bartlett

Estadísticos descriptivos

	Media	Desviación estándar	N de análisis
Matematicas	6,5000	2,61861	8
C Naturales	6,3750	1,99553	8
Francés	6,3750	2,61520	8
Latín	6,6250	1,99553	8

**Matriz de correlaciones<sup>a</sup>**

		Matematicas	C Naturales	Francés	Latín
Correlación	Matematicas	1,000	,943	,302	,560
	C Naturales	,943	1,000	,544	,758
	Francés	,302	,544	1,000	,934
	Latín	,560	,758	,934	1,000
Sig. (unilateral)	Matematicas		,000	,233	,074
	C Naturales	<b>,000</b>		,082	,015
	Francés	,233	,082		,000
	Latín	,074	<b>,015</b>	<b>,000</b>	
a. Determinante = ,001					

**p-valor**

**Determinante =.001 => var relacionadas**

**Matemáticas correlaciona significativamente con CN p-valor=0.000**

**Ciencias Nat correlaciona significativamente con Latín p-valor=0.015**

**Francés correlaciona significativamente con Latín p-valor=0.000**

*Las demás correlaciones No Significativas*

# Recordemos...

$$X_1 = a_{11}F_1 + a_{12}F_2 + \dots a_{1q}F_q + d_1U_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \dots a_{2q}F_q + d_2U_2$$

.....

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \dots a_{pq}F_q + d_p U_p$$

**$a_{ij}$  : saturaciones** de la variable  $X_i$  en el factor  $F_j$

$\mathbf{a_{ij}}^2$  es la **contribución** del Factor  $F_j$  a la variabilidad de la variable  $X_i$

**$d_j^2$**  es la contribución del factor único (**unicidad**)

La **suma de las contribuciones** de todos los factores **comunes**

$h_i^2 = a_{i1}^2 + \dots + a_{iq}^2$  se denomina **comunalidad**

# Donde aparecen...

$$X_1 = a_{11}F_1 + a_{12}F_2 + \dots a_{1q}F_q + d_1U_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \dots a_{2q}F_q + d_2U_2$$

.....

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \dots a_{pq}F_q + d_p U_p$$

**$a_{ij}$  : saturaciones. Están en la Matriz Factorial**

$\mathbf{d}_i^2$  es la contribución del factor único (**unicidad**). Están **en la diagonal de la matriz de Covarianzas anti-imagen**

**Comunalidad  $h_i^2 = a_{i1}^2 + \dots + a_{iq}^2$  están en la diagonal de la matriz reproducida**

- ✓ Descriptivos univariados
- ✓ Solución inicial

### Matriz de correlaciones

- ✓ Coeficientes
- ✓ Niveles de significación
- ✓ Determinante
- ✓ KMO y pr
- ✓ Reproducida
- ✓ Anti-Imagen

Reproducida: **Correlaciones entre los variables estimadas con el modelo factorial**

Covarianzas/Correlaciones **anti-imagen**.  
Covarianzas/correlaciones **parciales**,  
cambiadas de signo.

### Correlaciones reproducidas

		Matematicas	C Naturales	Francés	Latín
Correlación reproducida	Matematicas	,992 <sup>a</sup>	,952	,299	,562
	C Naturales	,952	,988 <sup>a</sup>	,544	,761
	Francés	,299	,544	,991 <sup>a</sup>	,945
	Latín	,562	,761	,945	,986 <sup>a</sup>
Residuo <sup>b</sup>	Matematicas		-,009	,004	-,001
	C Naturales	-,009		,000	-,003
	Francés	,004	,000		-,011
	Latín	-,001	-,003	-,011	

**Bajos**

Método de extracción: análisis de componentes principales.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Existen 0 (0,0%) residuos no redundantes con valores absolutos mayores que 0,05.

**Comunalidades proximas a 1**  
(en la diagonal de la matriz reproducida)

**Estimación de la unicidad**  
en la diagonal de la matriz de Covarianzas anti-imagen

### Matrices anti-imagen

		Matematicas	C Naturales	Francés	Latín
Covarianza anti-imagen	Matematicas	,047	-,037	,019	-,002
	C Naturales	-,037	,034	-,003	-,011
	Francés	,019	-,003	,056	-,042
	Latín	-,002	-,011	-,042	,039
Correlación anti-imagen	Matematicas	,570 <sup>a</sup>	-,918	,363	-,042
	C Naturales	-,918	,653 <sup>a</sup>	-,074	-,291
	Francés	,363	-,074	,570 <sup>a</sup>	-,903
	Latín	-,042	-,291	-,903	,661 <sup>a</sup>

a. Medidas de adecuación de muestreo (MSA)

**Medidas adecuación muestra**  
**Todas > 0.5 pero no muy altas**

Valores matriz *anti-imagen* bajos  
Los coeficientes de *correlación parciales* son muy bajos => existen factores comunes



*Analizando una y otra por separado...*

☒ Descriptivos univariados

☒ Solución inicial

### Matriz de correlaciones

☒ Coeficientes

☒ Niveles de significación ☒ Reproducida

☒ Determinante ☒ Anti-imagen

☒ KMO y prueba de esfericidad de Bartlett

Correlaciones entre los  
variables estimadas con el  
modelo factorial

### Correlaciones reproducidas

		Matemáticas	C Naturales	Francés	Latín
Correlación reproducida	Matemáticas	,992 <sup>a</sup>	,952	,299	,562
	C Naturales	,952	,988 <sup>a</sup>	,544	,761
	Francés	,299	,544	,991 <sup>a</sup>	,945
	Latín	,562	,761	,945	,986 <sup>a</sup>
Residuo <sup>b</sup>	Matemáticas		-,009	,004	-,001
	C Naturales	-,009		,000	-,003
	Francés	,004	,000		-,011
	Latín	-,001	-,003	-,011	

**Bajos**

Método de extracción: análisis de componentes principales.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Existen 0 (0,0%) residuos no redundantes con valores absolutos mayores que 0,05.

☐ **Comunalidades proximas a 1**  
(en la diagonal de la matriz reproducida)

- ☒ Descriptivos univariados
- ☒ Solución inicial

### Matriz de correlaciones

- ☒ Coeficientes ☐ Inverso
- ☒ Niveles de significación ☒ Reproducida
- ☒ Determinante ☒ Anti-imagen
- ☒ KMO y prueba de esfericidad de Bartlett

 **Estimación de la unicidad**  
 en la diagonal de la matriz de Covarianzas anti-imagen

Matrices anti-imagen

		Matematicas	C Naturales	Francés	Latín
Covarianza anti-imagen	Matematicas	<b>,047</b>	-,037	,019	-,002
	C Naturales	-,037	<b>,034</b>	-,003	-,011
	Francés	,019	-,003	<b>,056</b>	-,042
	Latín	-,002	-,011	-,042	<b>,039</b>
Correlación anti-imagen	Matematicas	<b>,570<sup>a</sup></b>	-,918	,363	-,042
	C Naturales	-,918	<b>,653<sup>a</sup></b>	-,074	-,291
	Francés	,363	-,074	<b>,570<sup>a</sup></b>	-,903
	Latín	-,042	-,291	-,903	<b>,661<sup>a</sup></b>

a. Medidas de adecuación de muestreo (MSA)

**Medidas adecuación muestra**  
 Todas > 0.5 pero no muy altas

Valores matriz *anti-imagen* bajos  
 Los coeficientes de *correlación parciales* son muy bajos => existen factores comunes

Covarianzas/Correlaciones **anti-imagen**.

Covarianzas/correlaciones **parciales**, cambiadas de signo.

*Volvemos a verlas juntas...*

☒ Descriptivos univariados

☒ Solución inicial

Matriz de correlaciones

☒ Coeficientes ☐ Inverso

☒ Niveles de significación ☒ Reproducida

☒ Determinante ☒ Anti-imagen

☒ KMO y prueba de esfericidad de Bartlett

Correlaciones reproducidas

		Matematicas	C Naturales	Francés	Latín
Correlación reproducida	Matematicas	<b>,992<sup>a</sup></b>	,952	,299	,562
	C Naturales	,952	<b>,988<sup>a</sup></b>	,544	,761
	Francés	,299	,544	<b>,991<sup>a</sup></b>	,945
	Latín	,562	,761	,945	<b>,986<sup>a</sup></b>
Residuo <sup>b</sup>	Matematicas		-,009	,004	-,001
	C Naturales	-,009		,000	-,003
	Francés	,004	,000		-,011
	Latín	-,001	-,003	-,011	

**Bajos**

Método de extracción: análisis de componentes principales.

a. Comunalidades reproducidas

b. Los residuos se calculan entre las correlaciones observadas y reproducidas. Existen 0 (0,0%) residuos no redundantes con valores absolutos mayores que 0,05.

**Comunalidades proximas a 1**  
(en la diagonal de la matriz reproducida)

**Estimación de la unicidad**

Matrices anti-imagen

		Matematicas	C Naturales	Francés	Latín
Covarianza anti-imagen	Matematicas	<b>,047</b>	-,037	,019	-,002
	C Naturales	-,037	<b>,034</b>	-,003	-,011
	Francés	,019	-,003	<b>,056</b>	-,042
	Latín	-,002	-,011	-,042	<b>,039</b>
Correlación anti-imagen	Matematicas	<b>,570<sup>a</sup></b>	-,918	,363	-,042
	C Naturales	-,918	<b>,653<sup>a</sup></b>	-,074	-,291
	Francés	,363	-,074	<b>,570<sup>a</sup></b>	-,903
	Latín	-,042	-,291	-,903	<b>,661<sup>a</sup></b>

a. Medidas de adecuación de muestreo (MSA)

**Medidas adecuación muestra**  
Todas > 0.5 pero no muy altas

**Valores matriz anti-imagen bajos**  
**Los coeficientes de correlación parciales son muy bajos => existen factores comunes**

Descriptivos...

☒ KMO y prueba de esfericidad de Bartlett

**Ho: Matriz de correlaciones = Matriz identidad**

(Unos en diagonal y ceros fuera de la diagonal)

Estructura esférica

$KMO \geq 0,75 \Rightarrow$  Bien

Kaiser-Meyer-Olkin para realizar un Análisis Factorial, proponen:  $KMO \geq 0,5 \Rightarrow$  Aceptable

$KMO < 0,5 \Rightarrow$  Inaceptable

## Prueba de KMO y Bartlett

Medida Kaiser–Meyer–Olkin de adecuación de muestreo		<b>,618</b> <b>&gt;0.5</b>
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	32,139
	gl	6
	Sig.	<b>,000</b>

**p-valor <0.05**

# Resultados aceptables

Si **p-valor <0.05**, se rechaza la esfericidad y eso significa que el AF puede ser adecuado

*Todo parece indicar que se puede encontrar  
esa estructura factorial ...*

**Busquemosla...**

☒ Descriptivos univariados

☒ Solución inicial

→ Permite obtener las **comunalidades iniciales** y los *autovalores de la matriz analizada* y los *porcentajes de varianza asociados a cada autovalor*.

Esta opción está activada por defecto.

Extracción...

✓ Componentes principales

Mínimos cuadrados no ponderados

Mínimos cuadrados generalizados

Máxima verosimilitud

Factorización de ejes principales

Análisis alfa

Análisis imagen



## Nº de FACTORES

## 2 Ejes factoriales

### Varianza total explicada

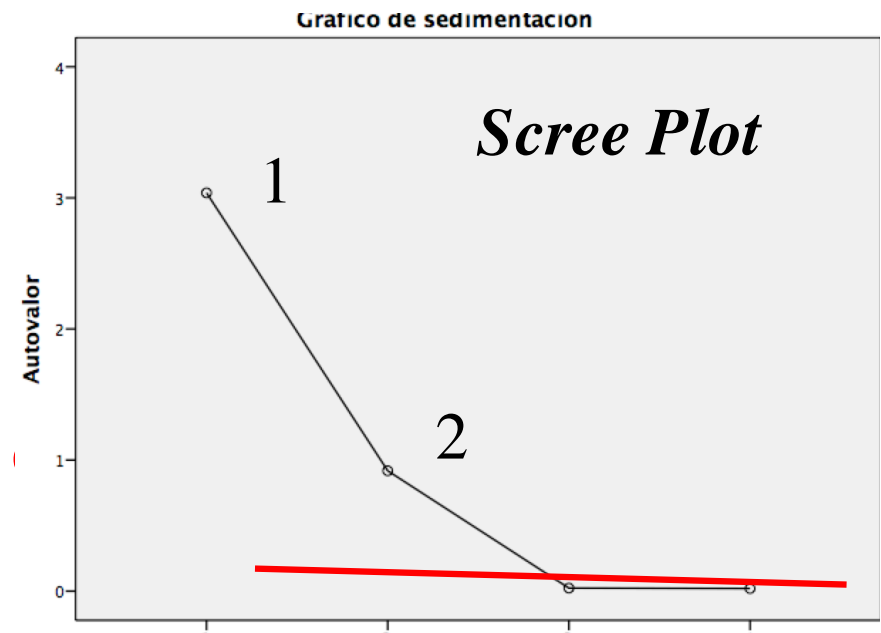
Componente	Autovalores iniciales		
	Autovalor Total	% de varianza	% acumulado
1	3,038	75,960	75.960
2	,918	22,960	98,920
3	,023	,583	99,503
4	,020	,497	100,000

Método de extracción: análisis de componentes principales.

Solo un autovalor >1

El primer Factor explica más del 75%

Los dos primeros absorben el 98.92%



MODELO

$a_{ij}$  saturaciones

$$X_i = a_{i1}F_1 + a_{i2}F_2 + d_iU_i$$

Solución SIN ROTAR

Matriz factorial<sup>a</sup>

	Factor	
	1	2
Matemáticas	,804	-,571
C Naturales	,934	-,337
Francés	,791	,582
Latín	,934	,335

Método de extracción: factorización de eje principal.

a. 2 factores extraídos. 6 iteraciones necesarias.

F1: Ciencias ???; F2: Letras ???

Estructura no clara

Análisis factorial: Extracción

Método: **Factorización de ejes principales**

Analizar

☒ Matriz de correlaciones

☐ Matriz de covarianzas

Visualización

☐ Solución factorial sin rotar

☒

Extraer

☐ Basado en autovalor

Autovalores mayores que: 1

☒ Número fijo de factores

Factores que extraer: 2

N.º máximo de iteraciones para conver

Ayuda

## Matriz de factor rotado<sup>a</sup>

	Factor	
	1	2
Matematicas	,973	,158
C Naturales	,901	,416
Francés	,154	,970
Latín	,429	,894

Método de extracción: factorización de eje principal.

Método de rotación: Varimax con normalización Kaiser.

# MODELO

$a_{ij}$  saturaciones

$$X_i = a_{i1}F_1 + a_{i2}F_2 + d_iU_i$$

## Solución SIN ROTAR

Matriz factorial<sup>a</sup>

	Factor	
	1	2
Matematicas	,804	-,571
C Naturales	,934	-,337
Francés	,791	,582
Latín	,934	,335

Método de extracción: factorización de eje principal.

a. 2 factores extraídos. 6 iteraciones necesarias.

## Solución ROTADA

Matriz de factores rotados<sup>a</sup>

	Factor	
	1	2
Matematicas	,973	,158
C Naturales	,901	,416
Francés	,154	,970
Latín	,429	,894

Método de extracción:

Factorización del eje principal.

Método de rotación:

Normalización Varimax con Kaiser.

F1: Ciencias; F2: Letras

## MODELO

$a_{ij}$  : saturaciones

$$X_i = a_{i1}F_1 + a_{i2}F_2 + d_i U_i$$

Matriz de factores rotados<sup>a</sup>

	Factor	
	1	2
Matemáticas	,973	,158
C Naturales	,901	,416
Francés	,154	,970
Latín	,429	,894

Método de extracción:

Factorización del eje principal.

Método de rotación:

Normalización Varimax con Kaiser.

$$\text{Matemáticas} = 0.973 C + 0.158 L + d_M U_M$$

$$CN = 0.901 C + 0.416 L + 0.034 U_M$$

$$\text{Francés} = 0.154 C + 0.970 L + 0.056 U_M$$

$$\text{Latín} = 0.429 C + 0.894 L + 0.039 U_M$$

## Estimación de la unicidad

Matrices anti-imagen

		Matemáticas	C Naturales	Francés	Latín
Covarianza anti-imagen	Matemáticas	0.047	-,037	,019	-,002
	C Naturales		,034	-,003	-,011
	Francés	,019	-,003	,056	-,042
	Latín	-,002	-,011	-,042	,039
Correlación anti-imagen	Matemáticas	,570 <sup>a</sup>	-,918	,363	-,042
	C Naturales	-,918	,653 <sup>a</sup>	-,074	-,291
	Francés	,363	-,074	,570 <sup>a</sup>	-,903
	Latín	-,042	-,291	-,903	,661 <sup>a</sup>

a. Medidas de adecuación de muestreo (MSA)

# *Resumiendo*

Toda la información interesante

**MODELO**  $X_i = a_{i1}F_1 + a_{i2}F_2 + d_iU_i$  Subíndice **i** denota la **Variable**  
 Subíndice **j** denota **Factor**

$a_{ij}$  : saturaciones

$a_{ij}^2$  es la **contribución** del Factor  $F_j$

$h_i^2 = a_{i1}^2 + \dots + a_{iq}^2$  se denomina **comunalidad**

La **suma de las contribuciones** de todos los factores comunes

$d_i^2$  es la contribución del factor único (**unicidad**)

$h_j^2 = a_{1j}^2 + \dots + a_{ij}^2$  Varianza explicada por el Factor j

	$a_{ij}$		$a_{ij}^2$		$h_i^2$	$h_i^2$	$d_i^2$
	<b>Saturaciones</b>		<b>Contribuciones</b>		<b>Comunalidad</b>	<b>Comunalidad</b>	<b>Unicidad</b>
	<i>Matriz Factor Principal rotada</i>		<i>Factor Principal rotada</i>		<i>Factor Principal rotada</i>	<i>Factor Principal Sin rotar</i>	<i>Diag Cov Antiimag</i>
	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 1</i>	<i>Factor 2</i>		<i>Suman 1</i>	
$X_1$ : Matemáticas	0.973	0.158	0.946	0.025	0.971	<b>0.953</b>	<b>0.047</b>
$X_2$ : CC Naturales	0.901	0.416	0.813	0.173	0.986	<b>0.966</b>	<b>0.034</b>
$X_3$ : Francés	0.154	0.970	0.024	0.941	0.965	<b>0.944</b>	<b>0.056</b>
$X_4$ : Latín	0.429	0.894	0.184	0.799	0.984	<b>0.961</b>	<b>0.039</b>
			<b>1.967</b>	<b>1.939</b>			

	<b>Saturaciones</b>		<b>Contribuciones</b>		<u>Comunalidad</u>	<u>Comunalidad Inicial (R²)</u>	<u>Unicidad Inicial</u>
	<b>Matriz Factor Principal rotada</b>		<b>Factor Principal rotada</b>		<i>Factor Principal rotada</i>	<i>Factor Principal Sin rotar</i>	<i>Diag Cov Antiimag</i>
	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 1</i>	<i>Factor 2</i>		<i>Suman 1</i>	
<b>X<sub>1</sub></b> : Matemáticas	0.973	0.158	0.946	0.025	0.971	<b>0.953</b>	<b>0.047</b>
<b>X<sub>2</sub></b> : CCNaturales	0.901	0.416	0.813	0.173	0.986	<b>0.966</b>	<b>0.034</b>
<b>X<sub>3</sub></b> : Francés	0.154	0.970	0.024	0.941	0.965	<b>0.944</b>	<b>0.056</b>
<b>X<sub>4</sub></b> : Latín	0.429	0.894	0.184	0.799	0.984	<b>0.961</b>	<b>0.039</b>
			<b>1.967</b>	<b>1.939</b>			

Varianza total explicada

Factor	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,038	75,960	75,960	3,016	75,398	75,398	1,967	49,175	49,175
2	,918	22,960	98,920	,890	22,250	97,648	1,939	48,473	97,648
3	,023	,583	99,503						
4	,020	,497	100,000						

Método de extracción: factorización de eje principal.



### Matriz de factor rotado<sup>a</sup>

	Factor	
	1	2
Matematicas	,973	,158
C Naturales	,901	,416
Francés	,154	,970
Latín	,429	,894

Método de extracción: factorización de eje principal.

Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 3 iteraciones.

### Matriz de factor rotado<sup>a</sup>

	Factor	
	1	2
Matematicas	,973	
C Naturales	,901	,416
Francés		,970
Latín	,429	,894

Método de extracción: factorización de eje principal.

Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 3 iteraciones.

☐ Excluir casos según

☐ Reemplazar por la m

#### Formato de presentación de los coeficientes

☐ Ordenados por tamaño

☒ Suprimir pequeños coeficientes

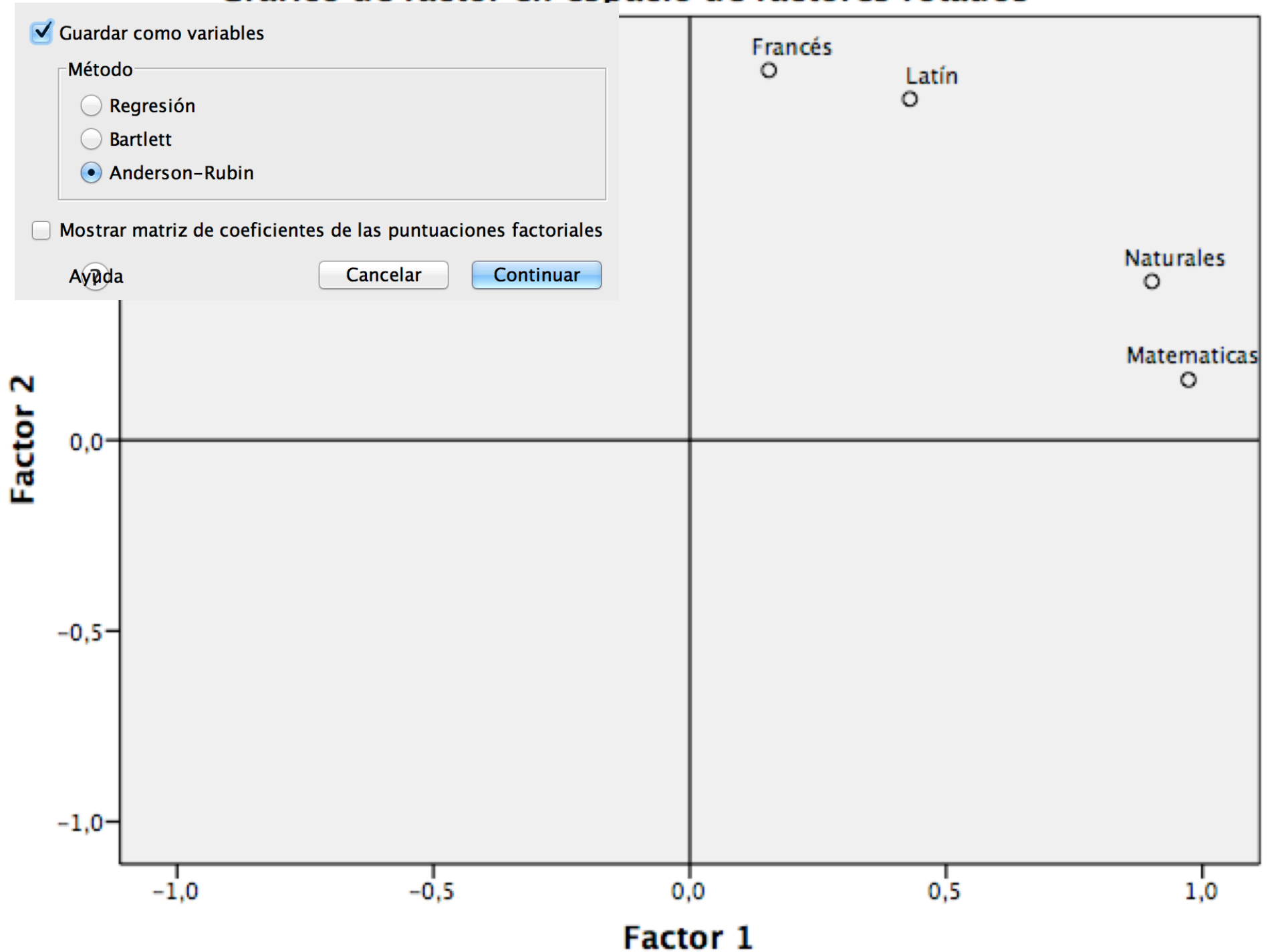
Valor absoluto bajo: ,30

Ayuda

Cancelar

Continuar

## Gráfico de factor en espacio de factores rotados



## Bibliografía

**CUADRAS, C.M.** (1996). *Métodos de Análisis Multivariante*, EUB, Barcelona.

**HAIR, J.F., ANDERSON, R.E., TATHAM, R.L. and BLACK, W.C.** (1998). *Multivariate Data Analysis*, Prentice Hall, New Jersey.

**JOLLIFE, I.T.** (1986). *Principal Component Analysis*, Springer-Verlag, New York.

**JOHNSON, D.E.** (1998). *Métodos Multivariados aplicados al análisis de datos*, Thomson Eds., México.