

4 Prácticas con SPSS

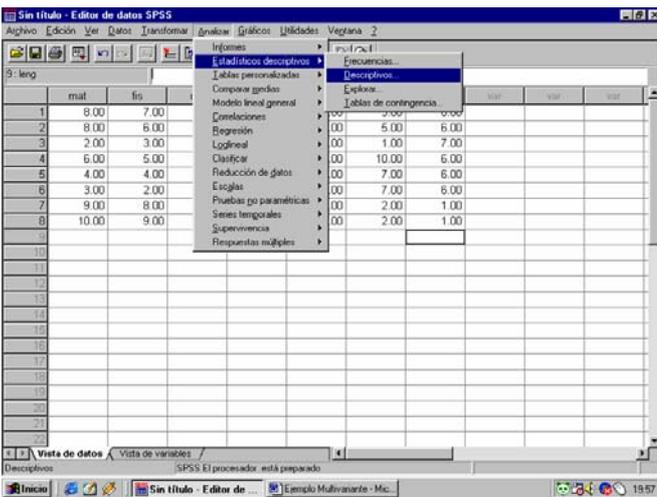
4.1 Análisis Multivariante mediante SPSS

Procedimiento Descriptivos

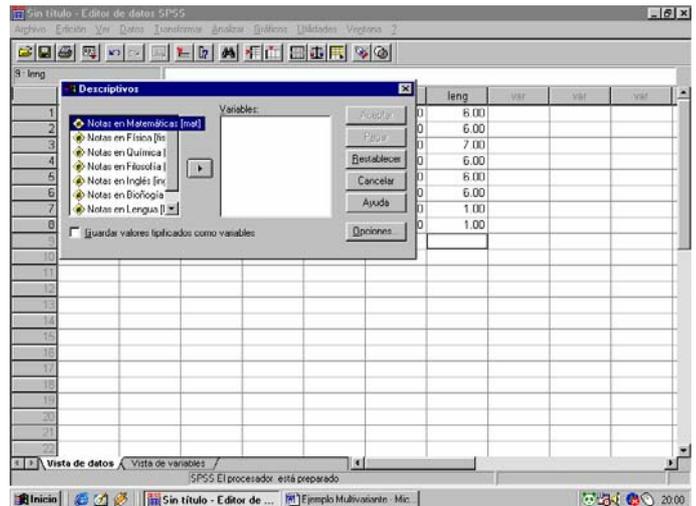
El procedimiento **Descriptivos** de SPSS muestra *estadísticos de resumen univariados* para *varias variables* en una única tabla y calcula *valores tipificados* (puntuaciones z). Los estadísticos que permite calcular son:

- tamaño de muestra,
- media,
- mínimo,
- máximo,
- desviación típica,
- varianza,
- rango,
- suma,
- error típico de la media,
- curtosis y
- asimetría con sus errores típicos.

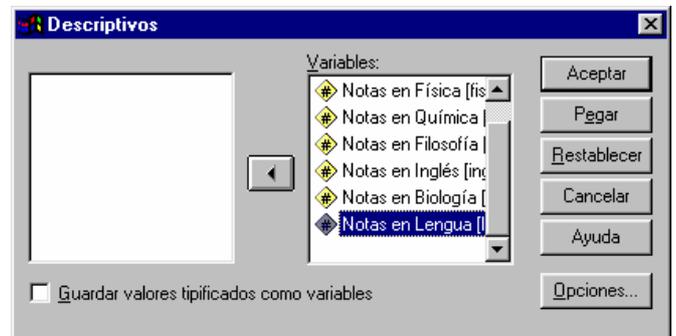
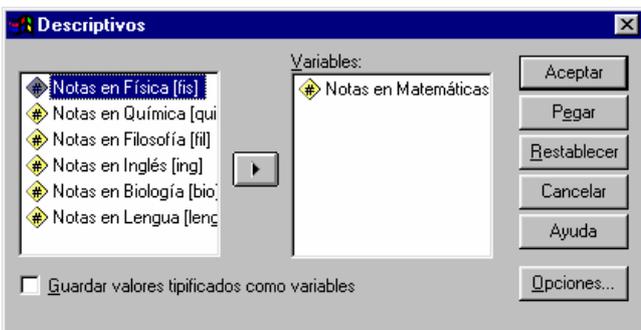
Para obtener estadísticos descriptivos :



Seleccione **Análisis / Estadísticos descriptivos / Descriptivos**.

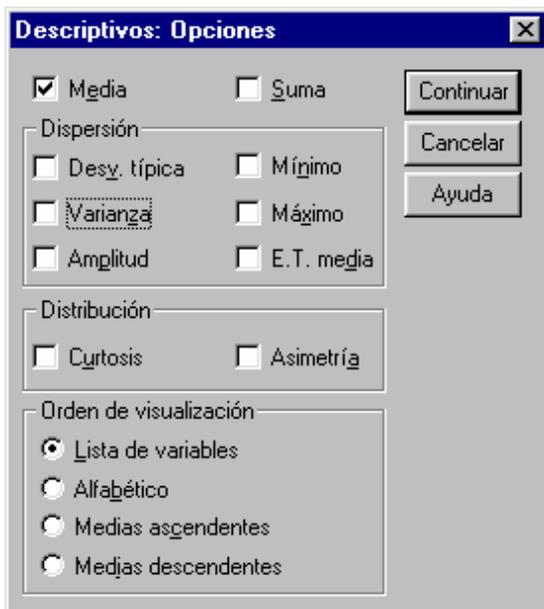


Seleccione una variable y pulse la flecha.

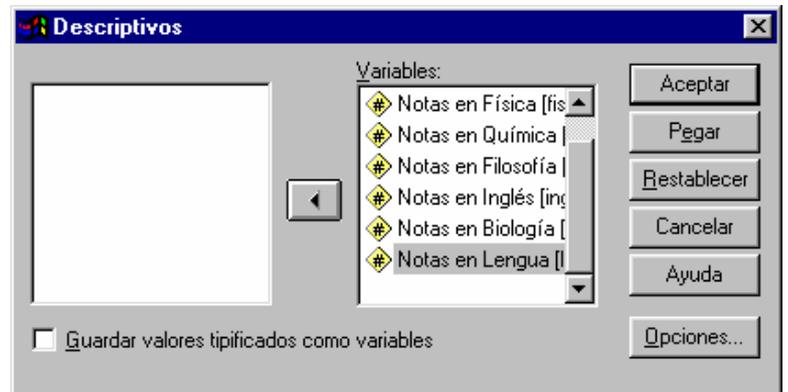


Repita el proceso con resto de las variables para seleccionarlas todas. Tiene la posibilidad de seleccionar la opción **Guardar valores tipificados como variables** para añadir al fichero de datos nuevas variables con los valores tipificados.

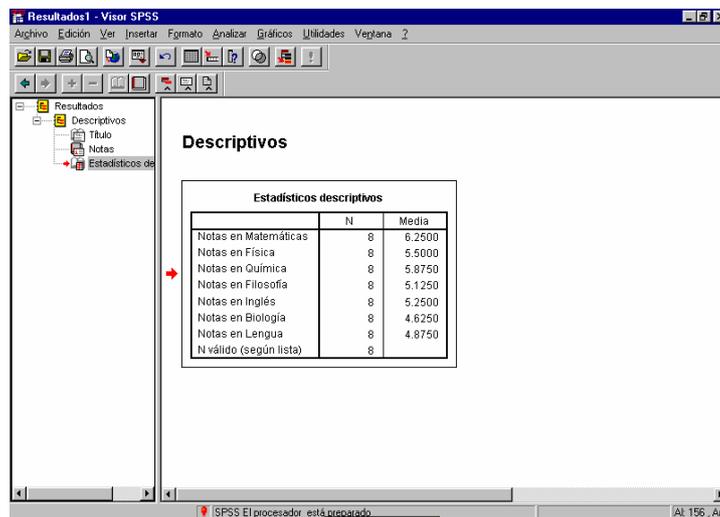
Pulse **Opciones** para seleccionar los estadísticos y el orden de visualización.



Active las casillas deseadas y pulse **Continuar**.



A continuación, en la ventana de diálogo **Descriptivos**, pulse **Aceptar**.



Procedimiento Correlaciones bivariadas

El procedimiento **Correlaciones bivariadas** de SPSS calcula el coeficiente de correlación de Pearson, la rho de Spearman y la tau-b de Kendall con sus niveles de significación.

Las correlaciones miden cómo están relacionadas las variables o los órdenes de los rangos. Antes de calcular el coeficiente de correlación conviene inspeccionar los datos para detectar valores atípicos (podrían producir resultados equívocos) y evidencias de una relación lineal.

El *coeficiente de correlación de Pearson* es una medida de asociación lineal. Si la relación entre dos variables, aunque perfecta, no es lineal, este coeficiente no será un estadístico adecuado para medir su asociación.

El que la *relación entre cada par de variables* sea o no *significativa* dependerá del correspondiente contraste realizado sobre el valor de r . La *hipótesis nula a contrastar* es

$$H_0: \text{“no hay relación entre las variables”} \quad \text{ó} \quad H_0: r = 0$$

Esta cuestión se resuelve analíticamente viendo si

$$|r| > \frac{2}{\sqrt{n}}$$

En caso afirmativo se dice que la relación es significativa y en caso negativo que no lo es.

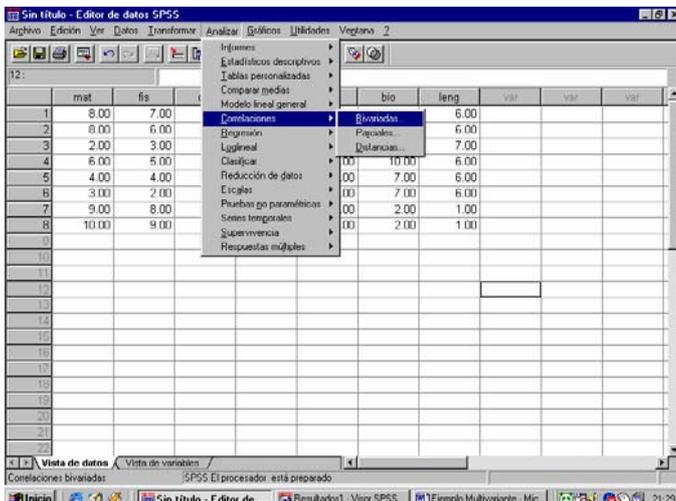
En cuanto a los estadísticos, para *una variable* calcula:

- número de casos sin valores perdidos,
- desviación típica,
- media.

y para *cada pareja de variables*:

- coeficiente de **correlación** de Pearson,
- rho de Spearman,
- tau-b de Kendall,
- productos cruzados de las desviaciones,
- **covarianzas**.

Para utilizar este procedimiento:



Seleccione **Analizar / Correlaciones / Bivariadas**.



Seleccione las variables deseadas para el estudio, al menos 2.

En el cuadro de diálogo de **Correlaciones bivariadas** se encuentran disponibles las siguientes opciones:

Coefficientes de correlación:

Para variables cuantitativas, normalmente distribuidas, seleccione el coeficiente de correlación de Pearson. Los coeficientes pueden estar entre -1 (relación negativa perfecta) y $+1$ (relación positiva perfecta). Un valor 0 indica que no existe relación lineal.

Si los datos no están normalmente distribuidos o tienen categorías ordenadas, delecione la tau-b de Kendall o de Spearman que miden la asociación entre órdenes de rangos.

Prueba de significación:

Se pueden seleccionar probabilidades bilaterales o unilaterales. Si se conoce de antemano la dirección de la asociación, se selecciona *Unilateral* y en caso contrario *Bilateral*.

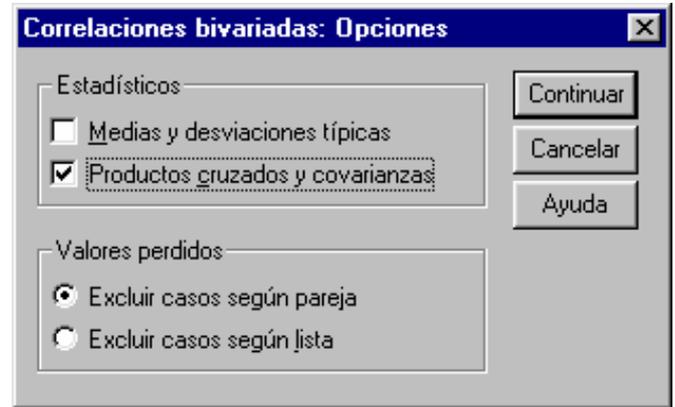
Marcar las correlaciones significativas:

Los coeficientes de correlación significativos al nivel 0.05 se identifican por medio de un sólo asterisco mientras que los significativos al nivel 0.01 mediante dos asteriscos.

Opciones:

Permite elegir estadísticos y manejar valores perdidos.

Para obtener la matriz de varianzas-covarianzas y la matriz de correlación:



Seleccione todas las variables y deje las opciones por defecto.

Pulse el botón **Opciones** y deje activada la casilla **Productos cruzados y covarianzas** para que, junto con la matriz de correlaciones, muestre la de covarianzas.

A continuación se obtiene la matriz de correlaciones y covarianzas con los coeficientes de significación de cada correlación.

En cuanto a las **correlaciones** decir que un *p-valor (sig.) pequeño* indica que se rechaza la hipótesis $r = 0$ (no hay relación lineal entre las variables) y, por tanto, *existe relación entre las variables*.

4.2 Análisis factorial mediante SPSS

Los cálculos necesarios para llevar a cabo un Análisis Factorial pueden efectuarse mediante diversos paquetes estadísticos. En lo que sigue se describe cómo hacerlos mediante SPSS.

Se comienza introduciendo los datos en un fichero o utilizando alguno disponible (Figura 1).

1:país	país	poblac	densidad	urbana	relig	espidaf	espidam	alfabet	in
1	Azerbaijón	7400	86.0	54	Musulma.	Esperanza de vida femenina		98	
2	Afganistán	20500	25.0	18	Musulma.	44	45	29	
3	Alemania	81200	227.0	85	Protest.	79	73	99	
4	Arabia Saudí	18000	7.7	77	Musulma.	70	66	62	
5	Argentina	33900	12.0	86	Católica	75	68	95	
6	Armenia	3700	126.0	68	Ortodoxa	75	68	98	
7	Australia	17800	2.3	85	Protest.	80	74	100	
8	Austria	8000	94.0	58	Católica	79	73	99	
9	Bahrein	600	828.0	83	Musulma.	74	71	77	
10	Bangladesh	125000	800.0	16	Musulma.	53	53	36	
11	Barbados	256	605.0	45	Protest.	78	73	99	
12	Bélgica	10100	329.0	96	Católica	79	73	99	
13	Bielorusia	10300	50.0	65	Ortodoxa	76	66	99	
14	Bolivia	7900	6.9	51	Católica	64	59	78	
15	Bosnia	4600	87.0	36	Musulma.	78	73	86	

Figura 1

En el menú se elige **Analizar / Reducción de Datos / Análisis Factorial**.



Figura 2

Seleccionadas las variables para el análisis (Figura 2) los botones *Descriptivos*, *Extracción*, *Rotación*, *Puntuaciones* y *Opciones* de la ventana de diálogo permiten introducir las especificaciones para el análisis.

▪ **Análisis factorial: Descriptivos**

El botón *Descriptivos* lleva a la pantalla en la que se establecen los estadísticos más relevantes relativos a las variables del análisis (Figura 3).

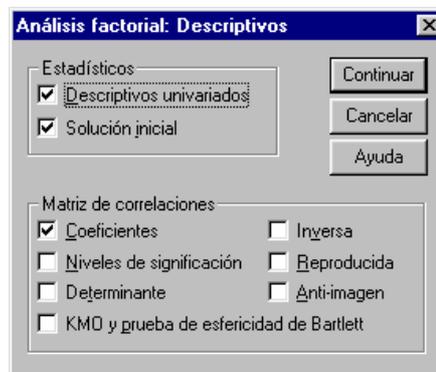


Figura 3

Estadísticos. Los estadísticos descriptivos univariados incluyen la media, desviación típica y el número de casos válidos para cada variable. La *solución inicial* muestra las comunales iniciales, los autovalores y el porcentaje de varianza explicada.

Matriz de correlaciones. Se dispone de las opciones: coeficientes, niveles de significación, determinante, inversa, reproducida, anti-imagen y KMO y prueba de esfericidad de Bartlett.

▪ **Análisis factorial: Extracción**

Lleva a la pantalla *Extracción* (Figura 4) cuyo cuadro *Método* permite especificar el método para la estimación del modelo factorial. Los métodos disponibles son:

- Componentes principales,
- Mínimos cuadrados no ponderados,
- Mínimos cuadrados generalizados,
- Máxima verosimilitud,
- Factorización de ejes principales,
- Factorización Alfa,
- Factorización imagen.

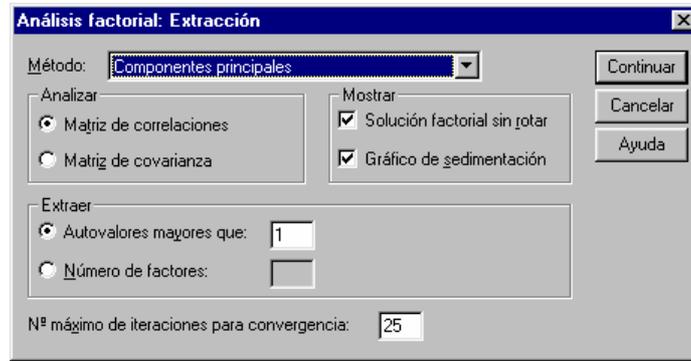


Figura 4

Analizar. Permite especificar una matriz de correlaciones o una matriz de covarianzas.

Extraer. Permite elegir todos los factores cuyos autovalores excedan un valor especificado (por defecto 1) o elegir un número específico de factores.

Mostrar. Permite pedir la solución factorial sin rotar y el gráfico de sedimentación de los autovalores.

Número máximo de iteraciones para convergencia. Se puede especificar el número máximo de pasos que el algoritmo sigue para estimar la solución.

▪ **Análisis Factorial: Rotación**

Lleva a la pantalla (Figura 5) cuyo cuadro **Método** permite seleccionar el método de rotación factorial. Los métodos disponibles son:

- Varimax,
- Equamax,
- Quartimax,
- Oblimin directo,
- Promax.

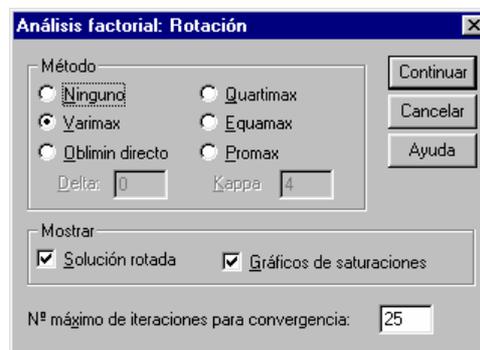


Figura 5

Mostrar. Permite incluir los resultados de la solución rotada, así como los gráficos de las saturaciones (coeficientes factoriales) para los dos o tres primeros factores.

Número máximo de iteraciones para convergencia. Se puede especificar el número máximo de pasos que el algoritmo puede seguir para llevar a cabo la rotación.

▪ **Análisis Factorial: Puntuaciones factoriales**

Permite especificar las opciones posibles en el cálculo y almacenamiento de las puntuaciones factoriales (Figura 6).

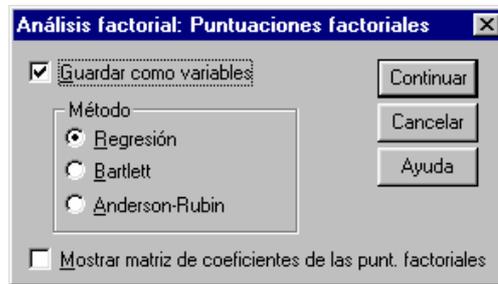


Figura 6

Guardar como variables. Se tiene la opción de guardar como variables las puntuaciones factoriales, creando una nueva variable para cada factor en la solución final. El cuadro *Método* permite seleccionar uno de los métodos alternativos para calcular las puntuaciones factoriales: Regresión, Bartlett, Anderson-Rubin (por defecto Regresión).

Mostrar matriz de coeficientes de las puntuaciones factoriales. Muestra los coeficientes por los que se multiplican las variables para obtener las puntuaciones factoriales. También muestra las correlaciones entre las puntuaciones factoriales.

▪ **Análisis Factorial: Opciones**

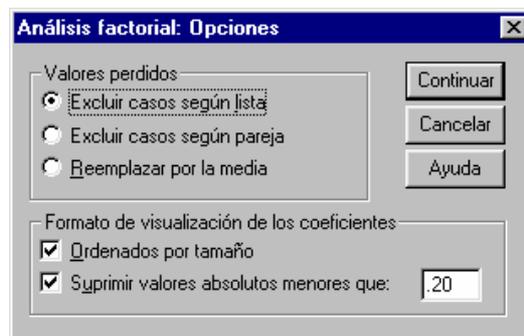


Figura 7

Valores perdidos. Permite especificar el tratamiento a dar a los valores perdidos. Las alternativas disponibles son:

- Excluir casos según lista, omite los casos que tienen valores perdidos en cualquiera de las variables seleccionadas para el análisis.
- Excluir casos según pareja, excluye los casos que tengan valores perdidos en cualquiera, o en ambas, de las variables de una pareja implicada en el cálculo de un estadístico específico.
- Reemplazar por la media, sustituye los valores perdidos por la media de la variable.

Formato de visualización de los coeficientes. Permite controlar aspectos de las matrices de resultados. Los coeficientes pueden ordenarse por tamaño y pueden suprimirse aquellos cuyos valores absolutos sean menores que el valor especificado.

En todas las ventanas de diálogo el botón **Continuar** permite pasar a la ventana de diálogo **Análisis factorial** para seguir fijando especificaciones. Una vez elegidas se pulsa **Aceptar** para obtener los resultados del Análisis Factorial.

4.3 Prácticas de Análisis Multivariante con SPSS

PRÁCTICA 1: EJEMPLO PRÁCTICO DE ANÁLISIS MULTIVARIANTE

PLANTEAMIENTO DEL PROBLEMA

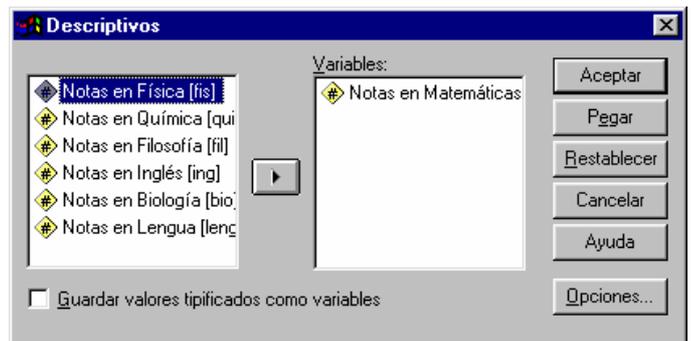
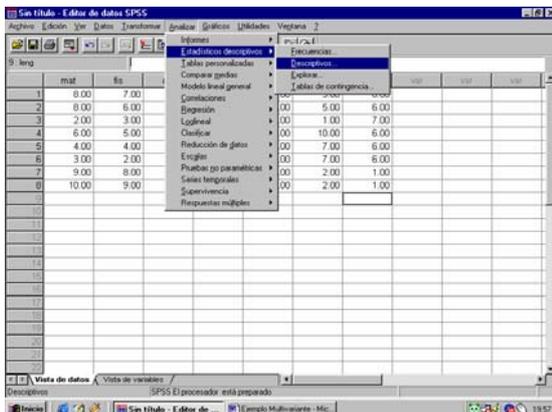
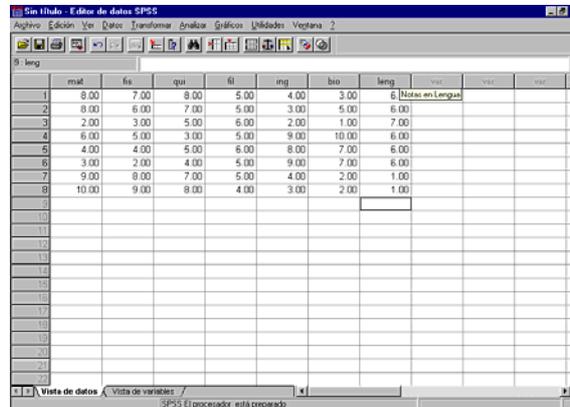
Se consideran las variables Mat, Fis, Qui, Fil, Ing, Bio y Leng correspondientes a las calificaciones en las materias Matemáticas, Física, Química, Filosofía, Inglés, Biología y Lengua, respectivamente. Se dispone de una muestra aleatoria de 8 individuos. Determínese el vector de medias, la matriz de varianzas y covarianzas y la matriz de correlaciones.

DATOS

	Mat	Fis	Qui	Fil	Ing	Bio	Leng	
	8	7	8	5	4	3	6	← Individuo 1
	8	6	7	5	3	5	6	← Individuo 2
	2	3	5	6	2	1	7	← Individuo 3
	6	5	3	5	9	10	6	← Individuo 4
	4	4	5	6	8	7	6	← Individuo 5
	3	2	4	5	9	7	6	← Individuo 6
	9	8	7	5	4	2	1	← Individuo 7
	10	9	8	4	3	2	1	← Individuo 8

RESOLUCIÓN DEL PROBLEMA MEDIANTE SPSS

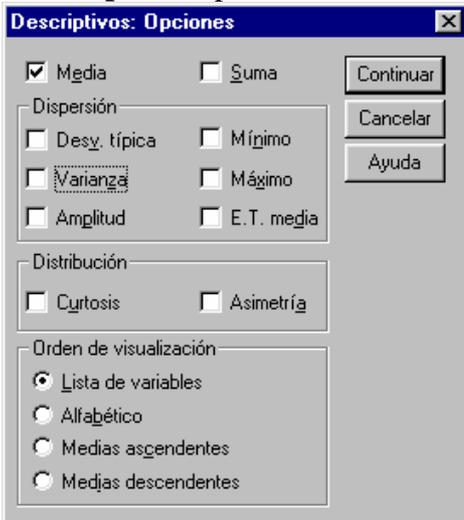
1. Se introducen los datos en un fichero.



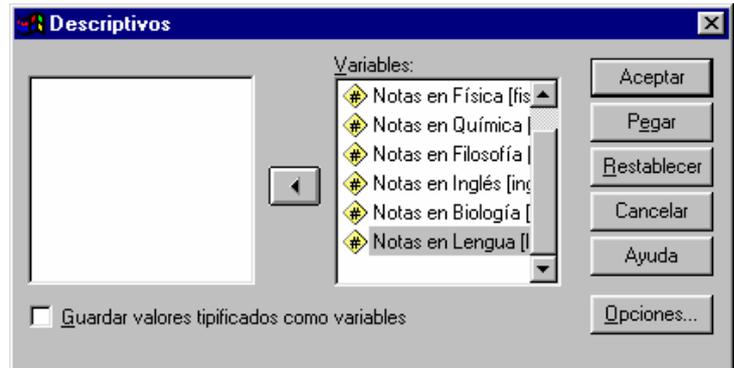
Seleccione **Analizar / Estadísticos descriptivos / Descriptivos**.

Seleccione las variables.

Pulse **Opciones** para seleccionar los estadísticos y el orden de visualización.

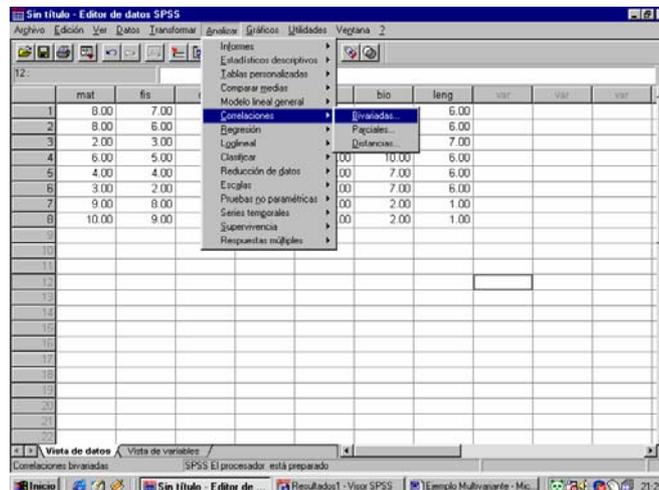


Active la casilla **Media** y pulse **Continuar**.

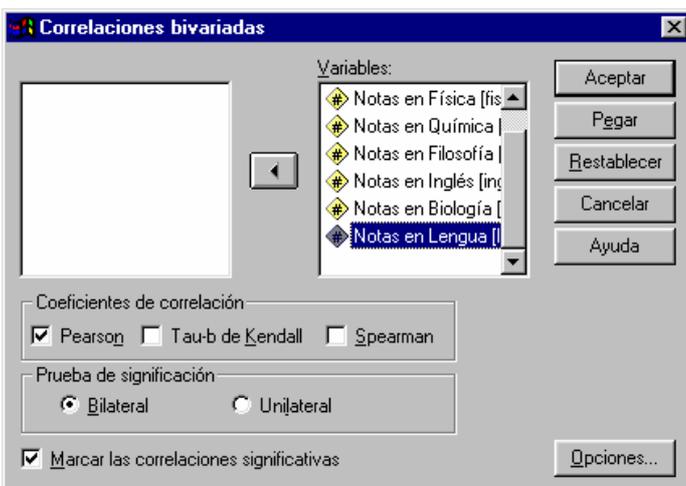


A continuación, en la ventana de diálogo **Descriptivos**, pulse **Aceptar**.

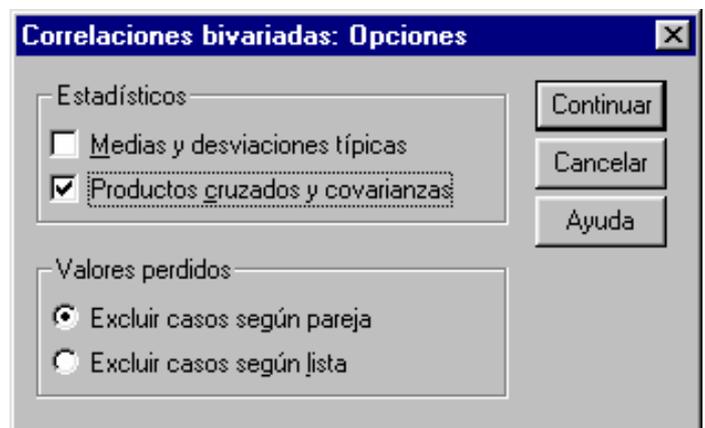
2. La matriz de varianzas-covarianzas y la matriz de correlaciones las vamos a obtener mediante Procedimiento Correlaciones bivariadas.



Seleccione **Analizar / Correlaciones / Bivariadas**.



Seleccione todas las variables y deje las opciones por defecto.



Pulse el botón **Opciones** y deje activada la casilla **Productos cruzados y covarianzas** para que junto con la matriz de correlaciones, muestre la de covarianzas.

ANÁLISIS DE LOS RESULTADOS

1. La salida de resultados tras aplicar el Procedimiento Descriptivos muestra el vector de medias:

The screenshot shows the SPSS 'Descriptivos' window. A table titled 'Estadísticos descriptivos' is displayed, showing the mean and number of cases (N) for each subject.

	N	Media
Notas en Matemáticas	8	6.2500
Notas en Física	8	5.5000
Notas en Química	8	5.8750
Notas en Filosofía	8	5.1250
Notas en Inglés	8	5.2500
Notas en Biología	8	4.6250
Notas en Lengua	8	4.8750
N válido (según lista)	8	

2. A continuación, aplicado el Procedimiento Correlaciones bivariadas se obtiene la matriz de varianzas-covarianzas y la matriz de correlaciones con los coeficientes de significación de cada correlación.

Correlaciones

		Notas en Matemáticas	Notas en Física	Notas en Química	Notas en Filosofía	Notas en Inglés	Notas en Biología	Notas en Lengua
Notas en Matemáticas	Correlación de Pearson	1.000	.964**	.748*	-.771*	-.388	-.278	-.733*
	Sig. (bilateral)	.	.000	.033	.025	.342	.504	.039
	Suma de cuadrados y productos cruzados	61.500	49.000	29.250	-10.250	-23.500	-18.250	-36.750
	Covarianza	8.786	7.000	4.179	-1.464	-3.357	-2.607	-5.250
	N	8	8	8	8	8	8	8
Notas en Física	Correlación de Pearson	.964**	1.000	.789*	-.683	-.500	-.434	-.784*
	Sig. (bilateral)	.000	.	.020	.062	.207	.283	.021
	Suma de cuadrados y productos cruzados	49.000	42.000	25.500	-7.500	-25.000	-23.500	-32.500
	Covarianza	7.000	6.000	3.643	-1.071	-3.571	-3.357	-4.643
	N	8	8	8	8	8	8	8
Notas en Química	Correlación de Pearson	.748*	.789*	1.000	-.458	-.747*	-.729*	-.537
	Sig. (bilateral)	.033	.020	.	.254	.033	.040	.170
	Suma de cuadrados y productos cruzados	29.250	25.500	24.875	-3.875	-28.750	-30.375	-17.125
	Covarianza	4.179	3.643	3.554	-.554	-4.107	-4.339	-2.446
	N	8	8	8	8	8	8	8
Notas en Filosofía	Correlación de Pearson	-.771*	-.683	-.458	1.000	.134	.097	.657
	Sig. (bilateral)	.025	.062	.254	.	.752	.819	.077
	Suma de cuadrados y productos cruzados	-10.250	-7.500	-3.875	2.875	1.750	1.375	7.125
	Covarianza	-1.464	-1.071	-.554	.411	.250	.196	1.018
	N	8	8	8	8	8	8	8
Notas en Inglés	Correlación de Pearson	-.388	-.500	-.747*	.134	1.000	.896**	.289
	Sig. (bilateral)	.342	.207	.033	.752	.	.003	.488
	Suma de cuadrados y productos cruzados	-23.500	-25.000	-28.750	1.750	59.500	57.750	14.250
	Covarianza	-3.357	-3.571	-4.107	.250	8.500	8.250	2.036
	N	8	8	8	8	8	8	8
Notas en Biología	Correlación de Pearson	-.278	-.434	-.729*	.097	.896**	1.000	.423
	Sig. (bilateral)	.504	.283	.040	.819	.003	.	.296
	Suma de cuadrados y productos cruzados	-18.250	-23.500	-30.375	1.375	57.750	69.875	22.625
	Covarianza	-2.607	-3.357	-4.339	.196	8.250	9.982	3.232
	N	8	8	8	8	8	8	8
Notas en Lengua	Correlación de Pearson	-.733*	-.784*	-.537	.657	.289	.423	1.000
	Sig. (bilateral)	.039	.021	.170	.077	.488	.296	.
	Suma de cuadrados y productos cruzados	-36.750	-32.500	-17.125	7.125	14.250	22.625	40.875
	Covarianza	-5.250	-4.643	-2.446	1.018	2.036	3.232	5.839
	N	8	8	8	8	8	8	8

** La correlación es significativa al nivel 0,01 (bilateral).

* La correlación es significativa al nivel 0,05 (bilateral).

En cuanto a las **correlaciones**, podemos decir lo siguiente. Un *p-valor (sig.)* pequeño indica que se rechaza la hipótesis $r=0$ (no hay relación lineal entre las variables) y, por tanto, *existe relación entre las variables*.

Esto ocurre, por ejemplo, entre las variables **Mat** y **Qui**:

$$r=0.748 \quad \text{y} \quad p\text{-valor}=0.033 < 0.05$$

Por el contrario, *no existe relación lineal* entre **Mat** y **Bio**:

$$r=-0.278 \quad \text{y} \quad p\text{-valor}=0.504 > 0.05$$

4.4 Prácticas de Análisis factorial con SPSS

PRÁCTICA 1: EJEMPLO PRÁCTICO DE ANÁLISIS FACTORIAL

PLANTEAMIENTO DEL PROBLEMA

Se desea construir una batería de pruebas que permita medir la memoria operativa, concepto que hace referencia al mantenimiento de la información activa durante un corto período de tiempo, incluyendo dos constructos teóricos fundamentales, el ejecutivo central y la memoria a corto plazo. Se decide incluir una prueba de dígitos hacia adelante (DA), otra de dígitos hacia atrás (DB), una prueba de capacidad de memoria (CM), y otra de bloques de Corsi (CB). Se aplican las pruebas a diez sujetos. Se va a efectuar un Análisis Factorial para reducir el conjunto de variables medidas a un conjunto menor de factores interpretando esos factores.

DATOS

Sujetos	DA	DB	CM	CB
1	7	5	6	6
2	6	5	6	5
3	8	6	7	7
4	7	6	5	5
5	8	4	5	7
6	8	5	5	7
7	9	5	5	8
8	7	4	5	6
9	6	5	5	6
10	7	5	6	5

RESOLUCIÓN DEL PROBLEMA MEDIANTE SPSS

1. Se introducen los datos en un fichero.

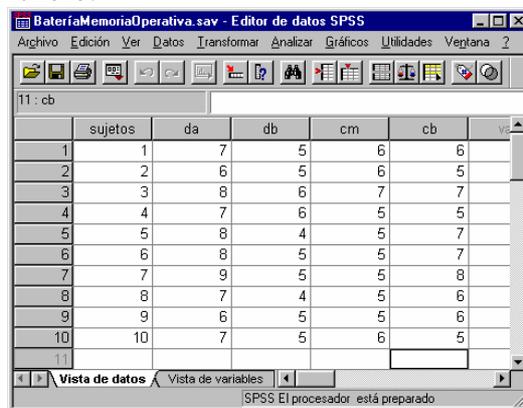


Figura 1

2. Se selecciona **Analizar/Reducción de Datos/Análisis Factorial**.

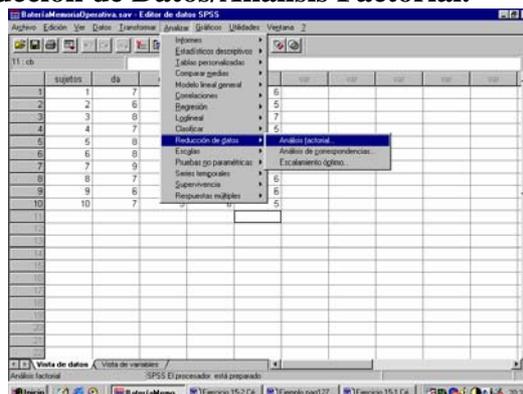


Figura 2

3. Se eligen las cuatro variables (todas excepto la ordinal).



Figura 3

4. Mediante los botones **Descriptivos**, **Extracción**, **Rotación** y **Puntuaciones** se especifican las características del análisis.

- **Descriptivos.** Se solicita la solución inicial. Para analizar las correlaciones se pide la matriz de coeficientes, el determinante, índice KMO y prueba de esfericidad de Barlett y la matriz anti-imagen.

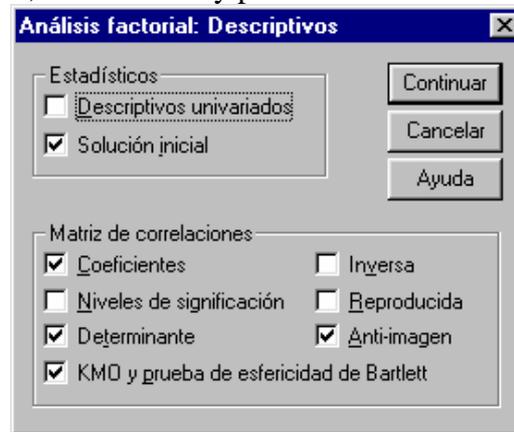


Figura 4

- **Extracción.** En el campo Método, se elige Ejes principales. En Extraer se activa Autovalores mayores que 1 (opción por defecto y la más recomendable). En el campo Mostrar se pide la Solución factorial sin rotar y el Gráfico de sedimentación.

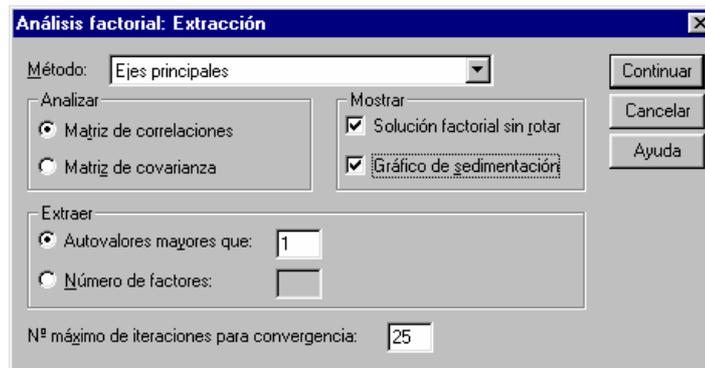


Figura 5

- **Rotación.** Se selecciona el Método Varimax. En Mostrar, Solución rotada y gráficos de saturaciones.

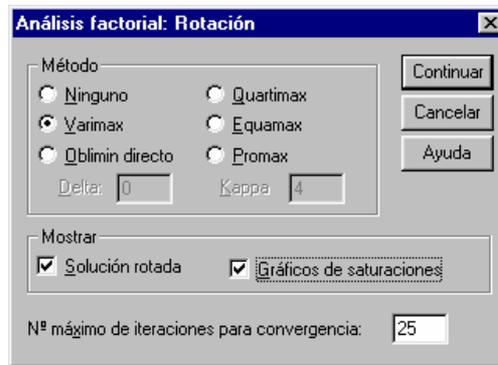


Figura 6

- **Puntuaciones.** Se selecciona Guardar como variables y Mostrar la matriz de coeficientes de las puntuaciones factoriales.

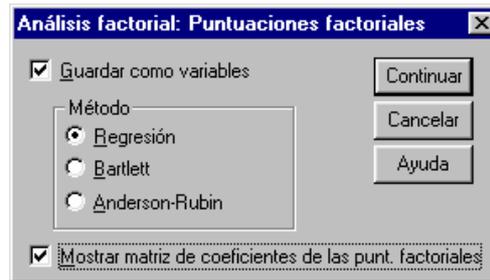


Figura 7

ANÁLISIS DE LOS RESULTADOS

El primer paso del análisis consiste en el *estudio de las correlaciones* entre todos los pares de variables consideradas.

En nuestro caso, un examen de la matriz de correlaciones (Figura 8) indica que hay variables que correlacionan muy alto entre sí (DA con CB), otras que correlacionan de forma moderada (por encima de 0.30, CM y DB) y que el resto de correlaciones son muy bajas. Esto parece indicar que serán dos los factores a extraer.

Matriz de correlaciones^a

		DA	DB	CM	CB
Correlación	DA	1.000	.000	-.083	.839
	DB	.000	1.000	.471	-.161
	CM	-.083	.471	1.000	-.152
	CB	.839	-.161	-.152	1.000

a. Determinante = .208

Figura 8

Cuando el número de variables es elevado (muy habitual) no es adecuada una inspección visual de la matriz de correlaciones, es necesario utilizar diversos índices que permitan saber si hay correlaciones altas en la matriz que permitan extraer factores.

El determinante de la matriz se emplea como índice del tamaño de las correlaciones. Cuando su valor es elevado, las correlaciones dentro de la matriz son bajas, mientras que un determinante bajo indica que hay algunas correlaciones altas en la matriz. En este ejemplo su valor es 0.208.

La prueba de esfericidad de Bartlett contrasta la hipótesis de que los elementos fuera de la diagonal principal (las correlaciones) de la matriz de correlaciones sean cero. En este caso el valor del estadístico es 10.733 con un p-valor $p=0.097$, próximo a la significatividad.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.483
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	10.733
	gl	6
	Sig.	.097

Figura 9

Otro índice es la medida de Kaiser-Meyer-Olkin que tiene en cuenta las correlaciones y las correlaciones parciales entre variables. Es aconsejable obtener valores grandes (más de 0.60) para que el análisis factorial pueda realizarse con garantías. En este ejemplo nos encontramos un valor mediocre, 0.483.

Matrices anti-imagen

		DA	DB	CM	CB
Covarianza anti-imagen	DA	.277	-.108	1.725E-02	-.232
	DB	-.108	.725	-.339	.114
	CM	1.725E-02	-.339	.771	6.573E-03
	CB	-.232	.114	6.573E-03	.270
Correlación anti-imagen	DA	.477 ^a	-.242	3.736E-02	-.849
	DB	-.242	.429 ^a	-.453	.258
	CM	3.736E-02	-.453	.549 ^a	1.442E-02
	CB	-.849	.258	1.442E-02	.489 ^a

a. Medida de adecuación muestral

Figura 10

Un indicador complementario al anterior es la matriz anti-imagen, que contiene el complementario de las correlaciones parciales. La interpretación es entonces simétrica a la anterior. En este ejemplo (Figura 10) encontramos una proporción alta de anti-imágenes bajas.

Una vez decidido si los datos son apropiados para una análisis factorial, si la decisión es positiva, puede pasarse a la *extracción de los factores*. En este ejemplo se ha solicitado el método de los ejes principales, método recursivo de computo semejante al de las componentes principales (es una generalización de esa técnica).

SPSS muestra en primer lugar la solución factorial inicial. En el caso de extracción por ejes principales, las comunalidades son las que muestra la Figura 11.

Comunalidades

	Inicial	Extracción
DA	.723	.876
DB	.275	.607
CM	.229	.367
CB	.730	.834

Método de extracción: Factorización de Ejes principales.

Figura 11

Las comunalidades representan la varianza de cada variable explicada por los factores o las componentes principales. La comunalidad de una variable es la suma de las cargas factoriales asociadas a ella elevadas al cuadrado. Su cómputo se realiza a partir de la matriz de cargas factoriales. En el caso de componentes principales, cuando se retienen todas las variables, la comunalidad es siempre 1. Cuando se emplean otros métodos de extracción, la *comunalidad inicial* es el coeficiente de correlación múltiple entre cada variable y todas las demás.

La *comunalidad final* suele tener mayor interés puesto que indica la variabilidad de cada variable que es explicada por el conjunto de factores retenidos

En este caso se observa, si se adopta una solución de dos factores, que se explica más del 80% de la varianza de las variables DA y CB (el 87.6% de la variable DA y el 83.4% de la variable CB), mientras que se explica el 60.7% de la variable DB y el 36.7% de CM.

La *elección del número de factores* a considerar en el modelo puede hacerse según diversos criterios.

Una aproximación razonable consiste en considerar el significado de los autovalores, indicadores de la proporción de varianza explicada. En este caso la varianza total, suma de los autovalores, es

$$\text{Varianza total} = 1.928 + 1.394 + 0.533 + 0.145 = 4$$

Así, el primer factor explica un

$$1.928/4 * 100 = 48.19\%$$

el segundo un 34.86%, el tercero un 13.32% y el cuarto un 3.63%, lo que indica que los dos primeros factores explican conjuntamente un 83.05% de la varianza total. Parece lógico seleccionar estos dos factores, reduciéndose la dimensionalidad del problema de 4 a 2.

Está generalizado el admitir que han de seleccionarse aquellos factores cuyo autovalor sea superior a 1. Obviamente esto conlleva una pérdida de capacidad explicativa, lo importante es que esa pérdida no sea significativa.

Todos estos datos, autovalores y varianzas explicadas antes y después de la extracción y después de la rotación se muestran en la Figura 12.

Varianza total explicada

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1.928	48.190	48.190	1.744	43.607	43.607	1.690	42.255	42.255
2	1.394	34.860	83.050	.940	23.498	67.105	.994	24.850	67.105
3	.533	13.321	96.371						
4	.145	3.629	100.000						

Método de extracción: Factorización de Ejes principales.

Figura 12

Un método alternativo para la selección del número de factores es el gráfico de sedimentación en el que se representan gráficamente los autovalores (eje de ordenadas) para cada factor (eje de abscisas). El análisis visual de este gráfico se centra en la búsqueda de un punto de inflexión en la gráfica, lo que habitualmente se produce con valores por debajo de 1.

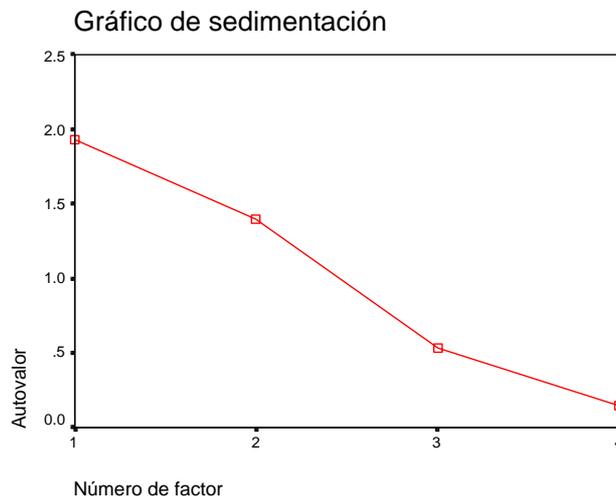


Figura 13

En este ejemplo la visión del gráfico de sedimentación justifica la selección de dos factores, algo que ya se había producido al fijar el criterio de autovalor en la unidad.

Los pesos de cada variable en cada factor aparecen bajo el epígrafe Matriz factorial (Figura 14).

Matriz factorial^a

	Factor	
	1	2
DA	.901	.254
DB	-.224	.746
CM	-.232	.560
CB	.910	7.492E-02

Método de extracción: Factorización del eje principal.

a. Se han intentado extraer 2 factores. Requidas más de 25 iteraciones. (Convergencia=4.345E-03). Se ha terminado la extracción.

Figura 14

Las soluciones factoriales rotadas facilitan la *interpretación de los factores*.

Cuando no se espera que los factores estén correlados se elegirá una rotación ortogonal en la que los ejes, después de la rotación, siguen siendo perpendiculares. El método Varimax es el más frecuentemente utilizado. Está orientado a la maximización de la varianza de los factores, lo que se traduce en que algunos factores tengan cargas altas y otras próximas a cero y de ahí que bastantes variables tengan saturaciones bajas en un factor. Los resultados del método Varimax son semejantes a los obtenidos por el método Quartimax. Con él se reduce el número de factores que son necesarios para explicar una variable. La rotación Equamax, mezcla de las dos anteriores, simplifica la interpretación puesto que reduce el número de variables que saturan alto en un factor y además reduce el número de factores necesarios para explicar una variable.

Las rotaciones oblicuas se utilizan cuando se espera que los factores estén correlacionados. Los ejes del espacio euclídeo no son ortogonales. Son poco recomendables para interpretar la estructura factorial (puede confundirse la interpretación de los factores). La mayoría de los investigadores prefieren las rotaciones no oblicuas.

En este ejemplo, la matriz de factores rotados con los pesos después de la rotación Varimax y la matriz de transformación empleada para realizar la rotación se muestra en las figuras 15 y 16.

Matriz de factores rotados^a

	Factor	
	1	2
DA	.936	1.195E-02
DB	-2.33E-02	.779
CM	-7.89E-02	.601
CB	.899	-.164

Método de extracción: Factorización del eje principal.
Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 3 iteraciones.

Figura 15

Matriz de transformación de los factores

Factor	1	2
1	.966	-.259
2	.259	.966

Método de extracción: Factorización del eje principal.
Método de rotación: Normalización Varimax con Kaiser.

Figura 16

Los valores de la matriz de factores rotados indican que las variables DA y CB tienen alta carga en el primer factor, mientras que DB y CM la tienen alta en el segundo. Se podría decir que las primeras dependen de un mecanismo diferente que las segundas. Examinando las tareas con detalle se concluiría que DA y CB implican un almacén a corto plazo y DB y CM la función ejecutiva de la memoria operativa.

La representación gráfica de las variables en el espacio factorial rotado aparece en la Figura 17. Las variables se sitúan más próximas al factor con el que están más correlacionadas.

Gráfico de saturaciones en espacio rotado

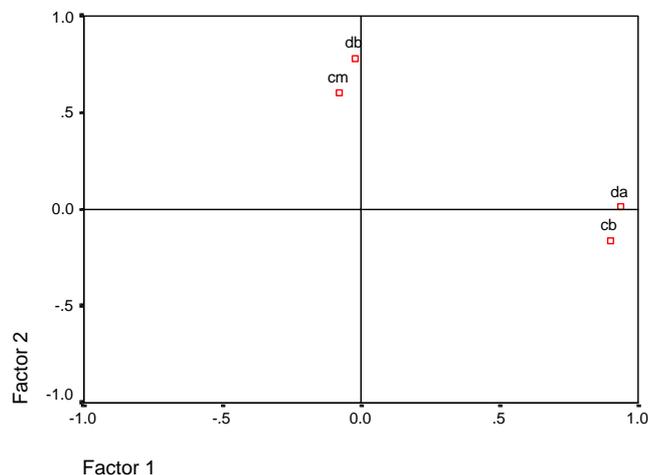


Figura 17

A veces interesa conocer las puntuaciones que obtendrían los sujetos si pudiesen ser medidos directamente en cada factor. Éstas sólo pueden conocerse de modo indirecto mediante las puntuaciones factoriales. En la Figura 18 se recogen los coeficientes para el cálculo de estas puntuaciones.

Matriz de coeficientes para el cálculo de las puntuaciones factoriales

	Factor	
	1	2
DA	.597	.180
DB	.035	.608
CM	.016	.303
CB	.406	-.171

Método de extracción: Factorización del eje principal.
Método de rotación: Normalización Varimax con Kaiser.

Método de puntuaciones factoriales: Regresión.

Figura 18

Matriz de covarianza de las puntuaciones factoriales

Factor	1	2
1	.921	-2.28E-02
2	-2.28E-02	.686

Método de extracción: Factorización del eje principal.
Método de rotación: Normalización Varimax con Kaiser.

Método de puntuaciones factoriales: Regresión.

Figura 19

Los valores de cada individuo en los dos nuevos factores aparecen en dos nuevas variables en el fichero de datos (Figura 20). El análisis de las puntuaciones factoriales de los sujetos proporciona una información que depende de la interpretación de los factores. Por ejemplo, si asociamos, como se ha indicado previamente, el primer factor a almacén a corto plazo, se tendría una medida de las diferencias entre los distintos sujetos en capacidad de almacén.

	sujetos	da	db	cm	cb	fac1_1	fac2_1	var
1	1	7	5	6	6	-.25601	.19027	
2	2	6	5	6	5	-1.27794	.16553	
3	3	8	6	7	7	.84049	1.55575	
4	4	7	6	5	5	-.61955	.83954	
5	5	8	4	5	7	.69133	-1.12573	
6	6	8	5	5	7	.74341	-.21347	
7	7	9	5	5	8	1.76533	-.18873	
8	8	7	4	5	6	-.33060	-1.15047	
9	9	6	5	5	6	-.90733	-.42845	
10	10	7	5	6	5	-.64913	.35576	

Figura 20

4.5 Prácticas de Análisis de componentes principales con SPSS

PRÁCTICA 1: EJEMPLO PRÁCTICO DE ANÁLISIS DE COMPONENTES PRINCIPALES

PLANTEAMIENTO DEL PROBLEMA

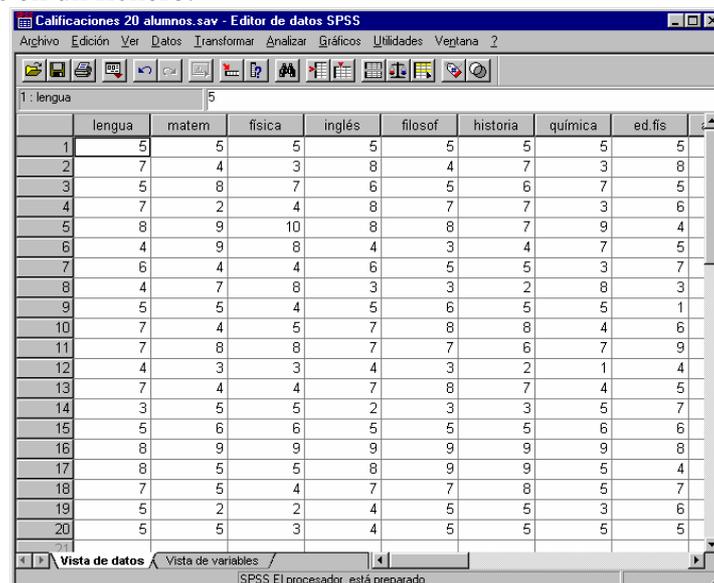
Se dispone de los datos de 20 alumnos de un grupo de Segundo Curso de Bachillerato relativos a las calificaciones en diversas asignaturas. Se pretende extraer un conjunto reducido de variables que resuman la información aportada por todas las variables, reduciendo la dimensionalidad del problema. Efectuar un Análisis de componentes principales.

DATOS

	Lengua	Matemáticas	Física	Inglés	Filosofía	Historia	Química	Ed.Física
1	5	5	5	5	5	5	5	5
2	7	4	3	8	4	7	3	8
3	5	8	7	6	5	6	7	5
4	7	2	4	8	7	7	3	6
5	8	9	10	8	8	7	9	4
6	4	9	8	4	3	4	7	5
7	6	4	4	6	5	5	3	7
8	4	7	8	3	3	2	8	3
9	5	5	4	5	6	5	5	1
10	7	4	5	7	8	8	4	6
11	7	8	8	7	7	6	7	9
12	4	3	3	4	3	2	1	4
13	7	4	4	7	8	7	4	5
14	3	5	5	2	3	3	5	7
15	5	6	6	5	5	5	6	6
16	8	9	9	9	9	9	9	8
16	8	5	5	8	9	9	5	4
18	7	5	4	7	7	8	5	7
19	5	2	2	4	5	5	3	6
20	5	5	3	4	5	5	5	5

RESOLUCIÓN DEL PROBLEMA MEDIANTE SPSS

1. Se introducen los datos en un fichero.



The screenshot shows the SPSS 'Calificaciones 20 alumnos.sav - Editor de datos SPSS' window. The data is entered in a grid format with columns for 'lengua', 'matem', 'fisica', 'inglés', 'filosof', 'historia', 'quimica', and 'ed.fis'. The rows correspond to the 20 students listed in the table above. The status bar at the bottom indicates 'Vista de datos' and 'SPSS El procesador está preparado'.

Figura 1

2. Se selecciona **Analizar/Reducción de Datos/Análisis Factorial**.

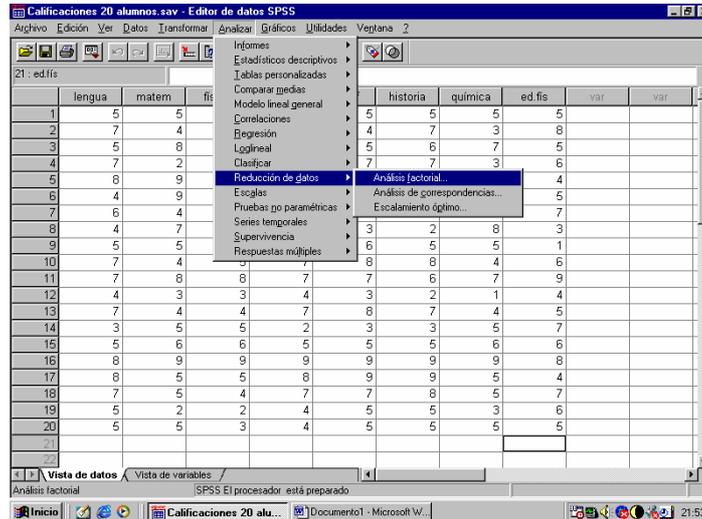


Figura 2

3. Se eligen las variables.



Figura 3

4. Mediante los botones **Descriptivos**, **Extracción**, **Rotación** y **Puntuaciones** se especifican las características del análisis.

- **Descriptivos.** Se solicita Descriptivos univariados y solución inicial.



Figura 4

- **Extracción.** En el campo Método, se elige Componentes principales. En Extraer se elige Autovalores mayores que 1 (opción por defecto y la más recomendable). En el campo Mostrar se pide la Solución factorial sin rotar y el Gráfico de sedimentación.

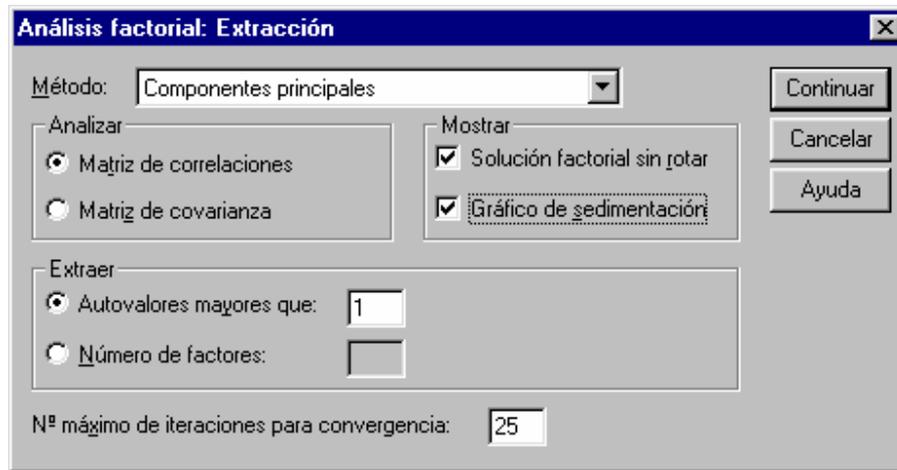


Figura 5

- **Rotación.** Se selecciona el Método Varimax. En Mostrar, Solución rotada y Gráfico de saturaciones.

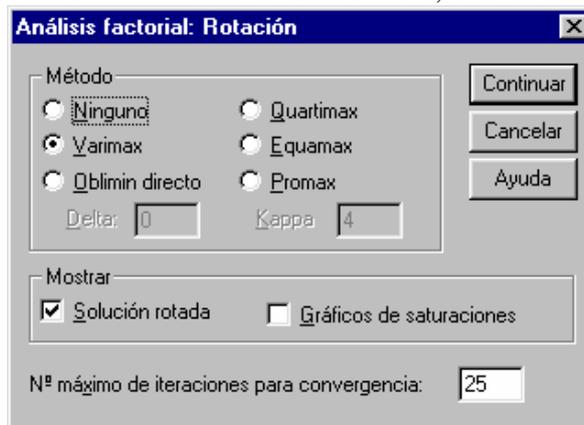


Figura 6

- **Puntuaciones.** Se selecciona Guardar como variables y Mostrar la matriz de coeficientes de las puntuaciones factoriales.

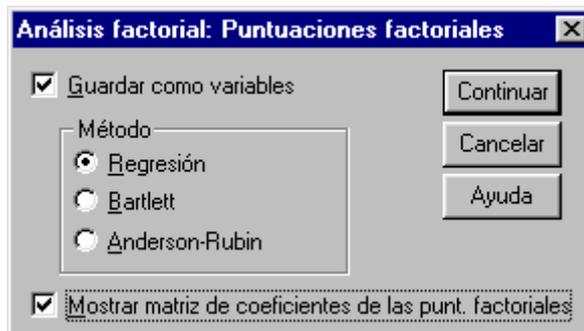


Figura 7

ANÁLISIS DE LOS RESULTADOS

En primer lugar aparecen los estadísticos descriptivos univariantes asociados a cada una de las variables consideradas para el análisis, concretamente media, desviación típica y tamaño de la muestra. La suma de las varianzas de cada variable representa la inercia total de la nube de puntos que habrá de ser explicada en la mayor medida con el modelo de análisis de componentes principales que se especifica.

Estadísticos descriptivos

	Media	Desviación típica	N del análisis
LENGUA	5.85	1.53	20
MATEMÁTICAS	5.45	2.21	20
FÍSICA	5.35	2.25	20
INGLÉS	5.85	1.95	20
FILOSOFÍA	5.75	2.02	20
HISTORIA	5.75	2.05	20
QUÍMICA	5.20	2.14	20
EDUCACIÓN FÍSICA	5.55	1.88	20

Figura 8

En la Figura 9 se muestra la tabla de las comunalidades asociada al análisis. Las comunalidades iniciales representan la varianza de cada variable explicada por todas las componentes principales (antes de la extracción el número de componentes coincide con el número de variables), por tanto valen siempre 1.

Comunalidades

	Inicial	Extracción
LENGUA	1.000	.957
MATEMÁTICAS	1.000	.948
FÍSICA	1.000	.941
INGLÉS	1.000	.920
FILOSOFÍA	1.000	.841
HISTORIA	1.000	.929
QUÍMICA	1.000	.948
EDUCACIÓN FÍSICA	1.000	.146

Método de extracción: Análisis de Componentes principales.

Figura 9

La comunalidad final indica la variabilidad de cada variable que es explicada por el conjunto de componentes retenidas por el modelo. Se puede observar que, salvo Educación Física (sólo es explicada en un 14.6%), las variables quedan bien explicadas por el modelo (Filosofía en un 84.10%, el resto en más de un 90%).

La Figura 10 contiene la tabla de valores propios y las inercias explicadas por las componentes, iniciales en el primer bloque, después de la extracción en el segundo, y después de la rotación en el tercero.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3.997	49.962	49.962	3.997	49.962	49.962	3.766	47.077	47.077
2	2.633	32.917	82.879	2.633	32.917	82.879	2.864	35.802	82.879
3	.930	11.626	94.505						
4	.201	2.511	97.016						
5	.128	1.605	98.621						
6	.053	.659	99.280						
7	.041	.515	99.794						
8	.016	.206	100.000						

Método de extracción: Análisis de Componentes principales.

Figura 10

Se observa que las dos primeras componentes son las que más peso tienen a la hora de explicar los datos. La primera explica un 49.962% de la inercia total de la nube de puntos y la segunda un 32.917% con lo que en total explican un 82.879%.

Teniendo en cuenta los porcentajes de variabilidad explicados por cada una de las componentes no seleccionadas para el modelo, se podría decidir incluir alguno más si se deseara conseguir una representación de la inercia total mayor que la conseguida con dos componentes (basta seleccionar un número fijo de componentes en Extracción).

El gráfico de sedimentación (Figura 11) justifica la elección de 2 componentes principales (punto de inflexión –comienzo de la zona de “sedimentación” de la curva- tras las 2 primeras componentes) para el modelo.

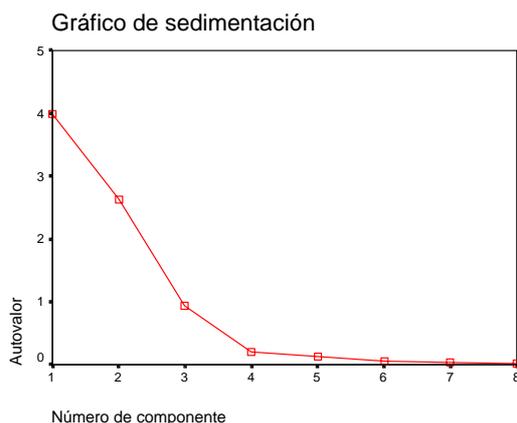


Figura 11

Los pesos de cada variable en cada componente antes de la rotación aparecen bajo el epígrafe Matriz de componentes (Figura 12). La suma de los cuadrados de los valores de cada columna coincide con el autovalor correspondiente a la componente asociada y la suma de los cuadrados de los valores de cada fila coinciden con la comunalidad de cada variable.

Estos coeficientes corresponden a los coeficientes de correlación lineal de Pearson entre la componente y la variable implicadas. Se observa que las variables Lengua, Inglés, Filosofía e Historia están relacionadas en mayor medida con la primera componente y las variables Matemáticas, Física y Química lo están con la segunda componente.

Matriz de componentes^a

	Componente	
	1	2
LENGUA	.925	-.320
MATEMÁTICAS	.405	.885
FÍSICA	.479	.844
INGLÉS	.910	-.302
FILOSOFÍA	.883	-.247
HISTORIA	.900	-.344
QUÍMICA	.470	.852
EDUCACIÓN FÍSICA	.329	-.194

Método de extracción: Análisis de componentes principales.

a. 2 componentes extraídos

Figura 12

Aunque en este caso parece clara la relación de cada componente con las variables, se ha solicitado una rotación varimax que identifique de forma más precisa esta relación (en el sentido de seguir el principio de estructura simple) si es posible.

Matriz de componentes rotados

	Componente	
	1	2
LENGUA	.974	8.919E-02
MATEMÁTICAS	4.657E-03	.974
FÍSICA	8.906E-02	.966
INGLÉS	.954	9.904E-02
FILOSOFÍA	.907	.138
HISTORIA	.962	5.649E-02
QUÍMICA	7.763E-02	.970
EDUCACIÓN FÍSICA	.380	-4.11E-02

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 3 iteraciones.

Figura 13

Matriz de transformación de las componentes

Componente	1	2
1	.911	.411
2	-.411	.911

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

Figura 14

Efectivamente, se corrobora lo que ya se había esbozado a la vista de la matriz factorial. Se observa que las variables Lengua, Inglés, Filosofía e Historia están altamente relacionadas con la primera componente, siendo prácticamente nula la correlación que presentan con la segunda componente. Las variables Matemáticas, Física y Química están relacionadas con la segunda componente y la correlación es casi cero con la primera componente.

La representación gráfica de los pesos de las componentes después de la rotación es la que se muestra en la Figura 15. Las variables se sitúan más próximas a la componente con la que están más correlacionadas.

Gráfico de componentes en espacio rotado

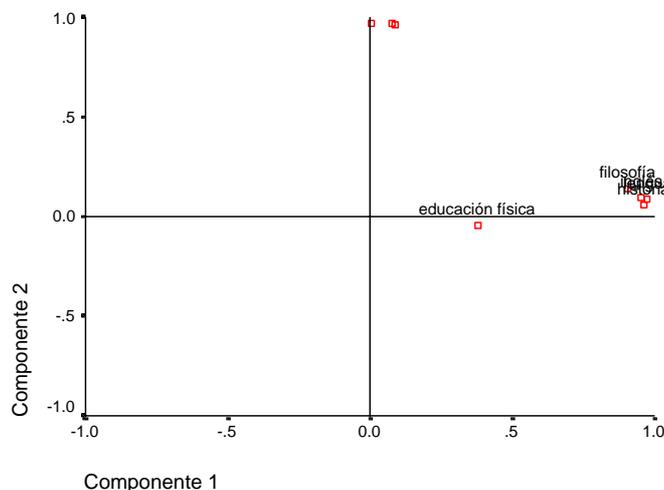


Figura 15

Las puntuaciones de cada individuo en las componentes principales aparecen en el fichero de datos como dos nuevas variables (Figura 18). En la Figura 16 se muestra la matriz de coeficientes para su cálculo.

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

	Componente	
	1	2
LENGUA	.231	-.121
MATEMÁTICAS	.101	.336
FÍSICA	.120	.320
INGLÉS	.228	-.115
FILOSOFÍA	.221	-.094
HISTORIA	.225	-.131
QUÍMICA	.118	.324
EDUCACIÓN FÍSICA	.082	-.074

Método de extracción: Análisis de componentes principales.

Figura 16

Matriz de covarianza de las puntuaciones de las componentes

Componente	1	2
1	1.000	.000
2	.000	1.000

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.
Puntuaciones de componentes.

Figura 17

	cpfac1_1	cpfac2_1	var	var	var	var	var	var
1	-.45502	-.12515						
2	.64490	-1.02447						
3	-.30499	.95193						
4	.92003	-1.14450						
5	.81773	1.85550						
6	-1.25724	1.30252						
7	.01319	-.80556						
8	-1.72343	1.21486						
9	-.55094	-.20318						
10	.97088	-.52267						
11	.61787	1.01227						
12	-1.31827	-1.31715						
13	.79832	-.64344						
14	-1.56483	-.10303						
15	-.44190	.32629						
16	1.55440	1.60461						
17	1.37147	-.19555						
18	.88552	-.37670						
19	-.41237	-1.38424						
20	-.56531	-.42234						

Figura 18