

Andrés Prieto

1004776434

Juan Pablo Prada Ortiz

1088030929

**TALLER ANÁLISIS MULTIVARIADO**

**FACULTAD DE INGENIERÍAS**

**MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

**SEPTIEMBRE 26 DE 2023**

1. Se dispone de los datos de 22 alumnos de un grupo de clases y se tienen las siguientes calificaciones en las asignaturas nombradas. **(VALOR 2.0)**

	Español	Matemáticas	Física	Inglés	Filosofía	Historia	Química	Estadística
1	4	9	5	8	7	6	5	4
2	8	3	4	8	7	9	5	3
3	5	8	7	6	5	6	7	8
4	7	2	4	8	7	7	3	6
5	8	9	10	8	8	8	9	10
6	4	9	8	4	3	4	7	10
7	6	4	4	6	7	6	3	2
8	4	7	8	3	3	2	8	3
9	5	5	4	5	6	5	5	2
10	7	4	5	7	8	8	4	8
11	8	8	8	7	7	8	7	9
12	4	3	3	4	3	2	1	4
13	7	4	4	7	8	7	4	5
14	3	6	7	2	3	3	8	7
15	5	6	6	5	5	5	6	7
16	7	9	9	9	9	9	9	10
17	8	5	5	8	9	9	5	3
18	7	5	4	7	7	8	5	3
19	5	2	2	4	5	5	3	5
20	4	3	3	4	5	5	5	2
21	1	8	6	2	5	5	9	6
22	5	8	7	6	5	7	9	10

Se pide:

- a. Realizar un análisis de componentes principales que resuma la información aportada por las variables

## Estadísticos descriptivos

	Media	Desv. Desviación	N de análisis
Español	5,5455	1,89554	22
Matematicas	5,7727	2,44816	22
Fisica	5,5909	2,15272	22
Ingles	5,8182	2,08478	22
Filosofia	6,0000	1,92725	22
Historia	6,0909	2,13606	22
Quimica	5,7727	2,30799	22
Estadistica	5,7727	2,89387	22

### 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	3,566	44,571	44,571	3,566	44,571	44,571	3,547	44,337	44,337
2	3,241	40,516	85,087	3,241	40,516	85,087	3,260	40,750	85,087
3	,501	6,262	91,349						
4	,275	3,444	94,793						
5	,260	3,252	98,045						
6	,093	1,161	99,206						
7	,050	,622	99,828						
8	,014	,172	100,000						

Método de extracción: análisis de componentes principales.

Según los resultados, el sistema decidió que en dos componentes se puede explicar las variables del modelo en un 85%.

- b. ¿Si se realiza una rotación se acomodan mejor las variables? Comparar ambas matrices

### Matriz de componente<sup>a</sup>

	Componente	
	1	2
Español	,870	-,289
Matematicas	,175	,905
Fisica	,281	,917
Ingles	,934	-,192
Filosofia	,904	-,255
Historia	,946	-,113
Quimica	,128	,909
Estadistica	,316	,747

Método de extracción: análisis de componentes principales.

a. 2 componentes extraídos.

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

	Componente	
	1	2
Español	,914	-,071
Matematicas	-,048	,921
Fisica	,052	,957
Ingles	,952	,038
Filosofia	,939	-,031
Historia	,945	,118
Quimica	-,094	,913
Estadistica	,127	,801

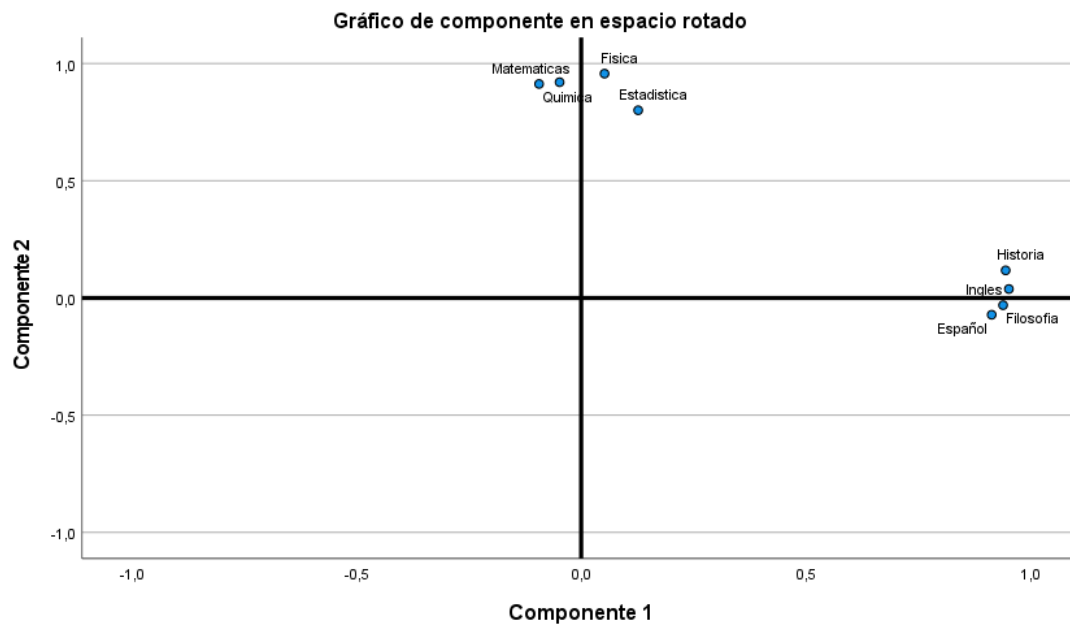
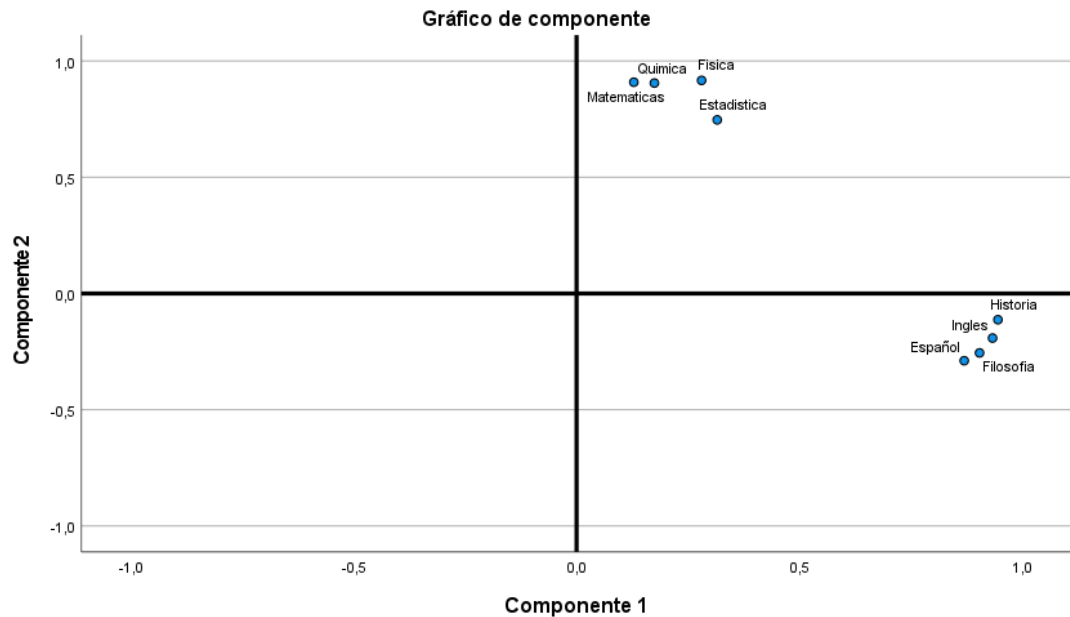
Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 3 iteraciones.

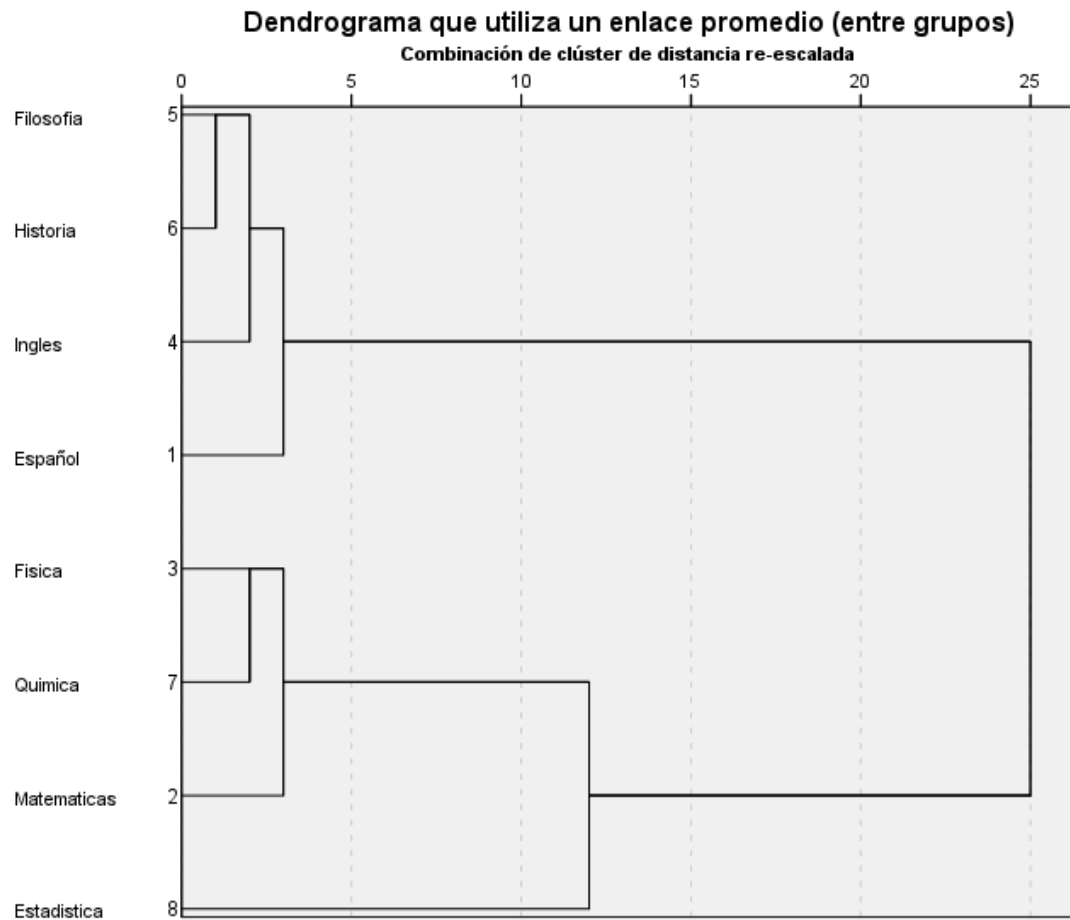
En efecto se puede observar que la relación de las variables es mucho mejor al hacer la rotación. A su vez, denota que en la componente 2 se encuentran las asignaturas en las que predomina el uso y tratamiento de número, mientras que en la componente 1 es todo lo contrario, se podría decir que son más literarias.

- c. Realizar un gráfico de dispersión de los casos una vez encontrado el número de componentes usando la mejor opción



A un nivel gráfico, se nota la afinidad de las variables con los componentes al realizar la rotación.

2. Usando el análisis clúster compare los resultados obtenidos del ejercicio 1 donde se aplicó componentes principales y haga un comparativo del dendograma con la matriz de componentes (**VALOR 1.0**)



En relación con la matriz de componentes, se tienen dos clústers los cuales destacan en que uno prima el manejo de los números y en el otro no.

3. En los últimos años en determinada universidad ha aumentado el porcentaje de estudiantes que abandonan sus estudios después del primer año. El año pasado, la universidad inició un programa voluntario de orientación para ayudar a los estudiantes de primer año a que se adapten a la vida del campus. Si se demuestra que el programa tiene resultados positivos, se considerará la posibilidad de que el programa sea obligatorio para todos los estudiantes de primer año. La administración de la universidad Collage supone que los estudiantes que tienen GPA bajo son los que tienen mayor probabilidad de abandonar los estudios al final del primer año. Con objeto de investigar la relación de estas variables con la permanencia de los estudiantes se tomó una muestra aleatoria de 100 estudiantes de primer año. Los datos se encuentran en el archivo Lakeland (**VALOR 1.5**).

Se pide:

- a.Cuál es la ecuación de regresión

Usando LINEST de LibreOffice Calc se puede obtener las siguientes tablas:

	Program X2	GPA X1	Return Y (Termino independiente)
<b>Parámetro</b>	0,301882425328202	0,321207977031571	-0,413475213265069
<b>Error estándar</b>	0,0891260491317056	0,0681661910577738	0,170142048549252
	0,403935591067764	0,37134036040035	#N/A
	32,8670456970931	97	#N/A
	9,06431466356063	13,3756853364394	#N/A

Por ende la ecuación correspondiente será:

$$Y = X_1 * 0,301882425328202 + X_2 * 0,321207977031571 - 0,413475213265069$$

Los valores de R y R<sup>2</sup> serían

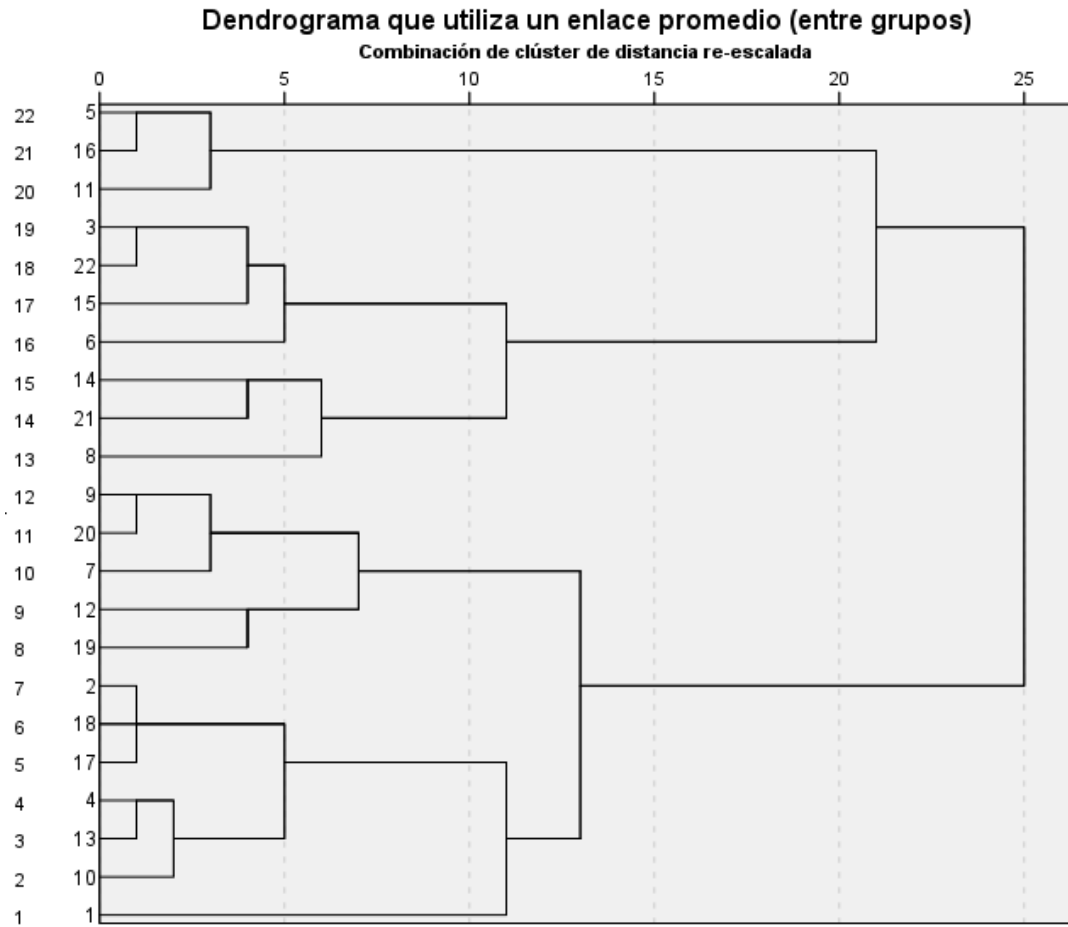
R	R <sup>2</sup>
0,636	0,404

b. Calcule el cociente de posibilidades para el programa de orientación

		Program X2	
Valor	0	36	-1,02857142857143
	1	64	-1,01587301587302
		<b>Cociente de posibilidades</b>	1,0125

Por tanto, se puede afirmar que (ya que el valor del cociente de posibilidades al ser mayor que 1) existe indicado que los estudiantes que participaron en el programa de orientación tenían una mayor probabilidad de éxito en comparación con aquellos que no lo hicieron.

4. Usando análisis clúster tomar los datos del punto 1 y acomodarlos en el número de clústers que se evidencien de acuerdo con el dendrograma y realizar una tabla de salida donde también se evidencie en cuál clúster quedaría cada estudiante (observación) **(VALOR 0.5)**



### Clúster de pertenencia

Caso	5 clústeres	4 clústeres	3 clústeres
1	1	1	1
2	2	1	1
3	3	2	2
4	2	1	1
5	4	3	3
6	3	2	2
7	5	4	1
8	3	2	2
9	5	4	1
10	2	1	1
11	4	3	3
12	5	4	1
13	2	1	1
14	3	2	2
15	3	2	2
16	4	3	3
17	2	1	1
18	2	1	1
19	5	4	1
20	5	4	1
21	3	2	2
22	3	2	2

En el dendograma se evidencia que el sistema agrupó a los estudiantes basados en su desempeño general, tomando a los estudiantes 5,11 y 16 como aquellos con el desempeño mas alto y a los estudiantes 12 y 19 con el mas bajo. Ahora, se decidió tomar de 3 a 5 clústers, si bien se bastaría con tener solo 3 los cuales serían:

1. Estudiantes con desempeño bajo
2. Estudiantes con desempeño medio
3. Estudiantes con desempeño alto

Decidimos dejarlo hasta 5 ya que se podría ser mas especifico a la hora de describir y agrupar los estudiantes.