

MEMORIA ANÁLISIS MULTIVARIANTE

CURSO 2023/2024



Profesor: Francisco Javier García Crespo

Asignatura: Análisis Multivariante

Autora: Noelia Cárdenas Salamanca

1. Tratamiento de la base de datos

1.1 Contenido

Para llevar a cabo este análisis, he elegido una base de datos de PISA que he obtenido del campus virtual. Esta base de datos contiene información recopilada a través de un cuestionario aplicado a estudiantes, en el cual se recoge información sobre aspectos relacionados con ellos, sus familias y hogares. Las preguntas más comunes abordan temas como el aprendizaje de idiomas en la escuela, sus opiniones sobre la lectura, su percepción de la calidad de vida, así como sus opiniones sobre el colegio, el horario escolar y el tiempo dedicado al aprendizaje.

ST012Q01TA_T: ¿Cuántas televisiones hay en la casa de los estudiantes?

ST012Q02TA_T: ¿Cuántos coches hay en la casa de los estudiantes?

ST012Q03TA_T: ¿Cuántas habitaciones con baño hay en la casa de los estudiantes?

ST012Q05NA_T: ¿Cuántos móviles con acceso a internet hay en la casa de los estudiantes?

ST012Q06NA_T: ¿Cuántos ordenadores hay en la casa de los estudiantes?

ST012Q07NA_T: ¿Cuántas tablets hay en la casa de los estudiantes?

ST012Q08NA_T: ¿Cuántos e-books hay en la casa de los estudiantes?

ST012Q09NA_T: ¿Cuántos instrumentos musicales hay en la casa de los estudiantes?

ST097Q01TA: ¿Con qué frecuencia durante las clases los estudiantes no escuchan al profesor?

ST097Q02TA: ¿Con qué frecuencia durante las clases hay ruido y desorden?

ST097Q03TA: ¿Con qué frecuencia durante las clases el profesor espera para que los estudiantes dejen de hablar?

ST097Q04TA: ¿Con qué frecuencia durante las clases los estudiantes no pueden trabajar bien?

ST097Q05TA: ¿Con qué frecuencia durante las clases los estudiantes no empiezan a trabajar cuando la clase empieza?

ST100Q01TA: ¿Con qué frecuencia durante las clases el profesor muestra un interés en el aprendizaje de los estudiantes?

ST100Q02TA: ¿Con qué frecuencia durante las clases el profesor da una ayuda extra cuando los estudiantes la necesitan?

ST100Q03TA: ¿Con qué frecuencia durante las clases el profesor ayuda a los estudiantes en su aprendizaje?

ST100Q04TA: ¿Con qué frecuencia durante las clases el profesor continúa explicando hasta que los estudiantes lo entiendan?

ST150Q01IA: En el último mes, ¿cuántas veces has tenido que leer en el colegio textos que contengan diagramas o mapas?

ST150Q02IA: En el último mes, ¿cuántas veces has tenido que leer en el colegio novelas o historias cortas de ficción?

ST150Q03IA: En el último mes, ¿cuántas veces has tenido que leer en el colegio textos que contengan tablas o gráficos?

ST150Q04HA: En el último mes, ¿cuántas veces has tenido que leer en el colegio textos digitales que contengan links.?

ST152Q05IA_T: En las clases, ¿con qué frecuencia el profesor ha motivado a los estudiantes a expresar sus opiniones de los textos?

ST152Q06IA_T: En las clases, ¿con qué frecuencia el profesor ha ayudado a relacionar las historias que leen los estudiantes con sus vidas?

ST152Q07IA_T: En las clases, ¿con qué frecuencia el profesor enseña a los estudiantes como está distribuida la información en los textos?

ST152Q08IA_T: En las clases, ¿con qué frecuencia el profesor propone preguntas que motivan a los estudiantes a ser más participativos?

El objetivo principal de este análisis es llevar a cabo un análisis factorial. Para lograrlo, será fundamental preparar la base de datos de manera que las conclusiones obtenidas sean lo más precisas posible. A partir de este punto, procederé a agrupar la información mediante factores, de modo que las variables queden representadas por estos. Una vez obtenidos estos factores, podremos validar el modelo mediante diversos métodos y extraer conclusiones.

1.2 Reducción y transformación de las variables

El primer paso que seguimos para preparar la base de datos es reducir el número de variables para disminuir el tamaño del problema.

Para seleccionar con qué variables nos quedamos, observamos cuales son las más relevantes, cuáles tienen más missings etc, y siguiendo estos criterios, nos quedamos con 41 variables.

Después de reducir la base de datos a las variables más importantes, procedo a recodificar dichas variables mediante **estandarización**. Este proceso garantiza que todas las variables sigan un criterio uniforme de valores, ordenados de menor a mayor.

En este último paso, he establecido que la respuesta 'siempre' o la opción más positiva se representa con el número 1, mientras que la respuesta 'nunca' o la opción más negativa se asigne al número 4. Este enfoque se aplica considerando que todas las variables seleccionadas contienen cuatro posibles respuestas.

1.3 Segmentación por países y tratamiento de valores faltantes (*missings*)

Este paso consiste en segmentar los datos por países, permitiéndonos así obtener conclusiones específicas sobre cada uno de ellos e identificar aquellas variables que presentan mayor cantidad de *missings*. Esta última observación se llevará a cabo mediante un análisis de valores perdidos, abarcando todas las variables seleccionadas.

Una vez realizado este análisis de *missings*, procedo a eliminar los países que presentan un exceso de ellos. En este contexto, considero como exceso aquellos países que contienen más del 10% de datos perdidos en un conjunto de variables o en diferentes conjuntos.

Los países que decido eliminar son:

- Baku 31, Australia 36, Bulgaria 100, Canada 124, Costa rica 188, Denmark 208, Dominican Republic 214, Georgia 268, Germany 276, Israel 376, Lebanon 422, México 484, Netherlands 528, Panamá 591, North macedonia 807, y Uruguay 858.

Ahora contamos con un número de países más reducido pero que aún sigue teniendo datos faltantes, pero en un porcentaje razonable para poder estimarlos.

Para ello, vuelvo a realizar un análisis de valores perdidos en el que selecciono las variables a estudiar y creo una nueva base de datos la cual no contiene ningún missing.

Ahora la base de datos está lista para llevar a cabo un análisis factorial. Antes de entrar en detalles, proporciono una descripción del análisis que vamos a realizar:

Mirando esta matriz, aprecio la presencia de cinco conjuntos diferentes donde hay una correlación muy fuerte entre las variables dentro del mismo conjunto. En contraste, la correlación es muy baja entre una variable de un conjunto y otra de un conjunto diferente. En resumen, podemos anticipar la presencia de cinco factores.

2.2 Índices KMO y Bartlett

A continuación, calculo el índice KMO y el test de esfericidad de Bartlett, ya que estos son los que nos van a indicar si tiene o no sentido aplicar el análisis factorial.

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		0,864
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	432914,1266
	gl	300
	Sig.	0,000

El **índice KMO** considera los coeficientes de correlación parcial entre pares de variables.

En nuestro caso observamos que dicho índice es 0,864 que es mayor que 0,5 lo cual implica que el valor de los coeficientes de correlación parcial entre nuestras variables es bajo y, por lo tanto, será viable el análisis factorial.

Por otra parte, el **contraste de esfericidad de Bartlett** se emplea para analizar el parecido de la matriz de correlaciones a la matriz identidad, de hecho, si la matriz de correlaciones fuera la identidad implicaría que la correlación entre las distintas variables es 0 y por tanto no se podría realizar una reducción factorial. En este caso tenemos un p-valor de 0 que al ser menor que 0,05 no nos proporciona evidencias suficientes para aceptar la hipótesis nula (matriz de correlaciones parecida a la identidad).

2.3 Interpretación de resultados

Claramente, podemos utilizar el Análisis Factorial como técnica para analizar nuestros datos y determinar el número de factores que representarán a las variables originales. En este proceso, emplearé el **método de componentes principales** para extraer estos factores. Aunque este método tiene la desventaja de generar estimaciones sesgadas de la matriz de cargas factoriales en casos de bajas comunalidades, pero su ventaja es que siempre proporcionará una solución.

- En primer lugar, obtengo la tabla de estadísticos descriptivos:

Estadísticos descriptivos			
	Media	Desv. Desviación	N de análisis
How many in your home:Televisions	1,95	0,836	489979
How many in your home:Cars	2,57	0,963	489979
How many in your home:Rooms with a bath or shower	2,50	0,840	489979
How many in your home:<Cell phones> with Internet access (e.g. smartphones)	1,38	0,785	489979
How many in your home:Computers (desktop computer, portable laptop, or notebook)	2,27	0,997	489979
How many in your home:<Tablet computers> (e.g. <iPad>, <BlackBerry PlayBook>)	2,84	1,043	489979
How many in your home:E-book readers (e.g. <Kindle>, <Kobo>, <Bookeen>)	3,67	0,682	489979
How many in your home:Musical instruments (e.g. guitar, piano)	2,99	1,074	489979

Esta tabla nos indica los estadísticos descriptivos univariantes asociados a cada una de las variables consideradas para el análisis como la media, desviación típica y el tamaño de la muestra.

- En segundo lugar, obtengo la tabla de comunalidades:

Comunalidades		
	Inicial	Extracción
How often during <test language lessons>: Students don't listen to what the teacher says.	1,000	0,642
How often during <test language lessons>: There is noise and disorder.	1,000	0,708
How often during <test language lessons>: The teacher waits long for students to quiet down.	1,000	0,701
How often during <test language lessons>: Students cannot work well.	1,000	0,619
How often during <test language lessons>: Students don't start working for a long time after the lesson begins.	1,000	0,643
How often during <test language lessons>: The teacher shows an interest in every student's learning.	1,000	0,656
How often during <test language lessons>: The teacher gives extra help when students need it.	1,000	0,741
How often during <test language lessons>: The teacher helps students with their learning.	1,000	0,772
How often during <test language lessons>: The teacher continues teaching until the students understands.	1,000	0,700

En esta tabla podemos ver las comunialidades iniciales, que representan la varianza de cada variable explicada por las componentes principales.

Debemos recordar que como antes de la extracción el número de componentes coincide con el de variables, es coherente que el valor sea 1.

Las comunialidades finales, nos indican la variabilidad que es explicada por el conjunto de componentes que hemos elegido mantener.

Por lo que se puede decir que, la mayoría de las variables quedan lo suficientemente bien explicadas ya que se encuentran por encima del 0,5.

Si observamos la tabla completa podremos ver cómo algunas variables tienen una comunialidad por debajo de 0,5 pero no debemos eliminarlas ya que eso haría que perdiéramos información sobre la totalidad del grupo al que pertenecen.

- En tercer lugar, obtengo la tabla de la varianza total explicada.

Varianza total explicada						
Componente		Autovalores iniciales		Sumas de cargas al cuadrado de la extracción		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,980	19,918	19,918	4,980	19,918	19,918
2	3,065	12,260	32,178	3,065	12,260	32,178
3	2,651	10,602	42,780	2,651	10,602	42,780
4	2,014	8,057	50,837	2,014	8,057	50,837
5	1,593	6,371	57,208	1,593	6,371	57,208
6	0,998	3,994	61,202			

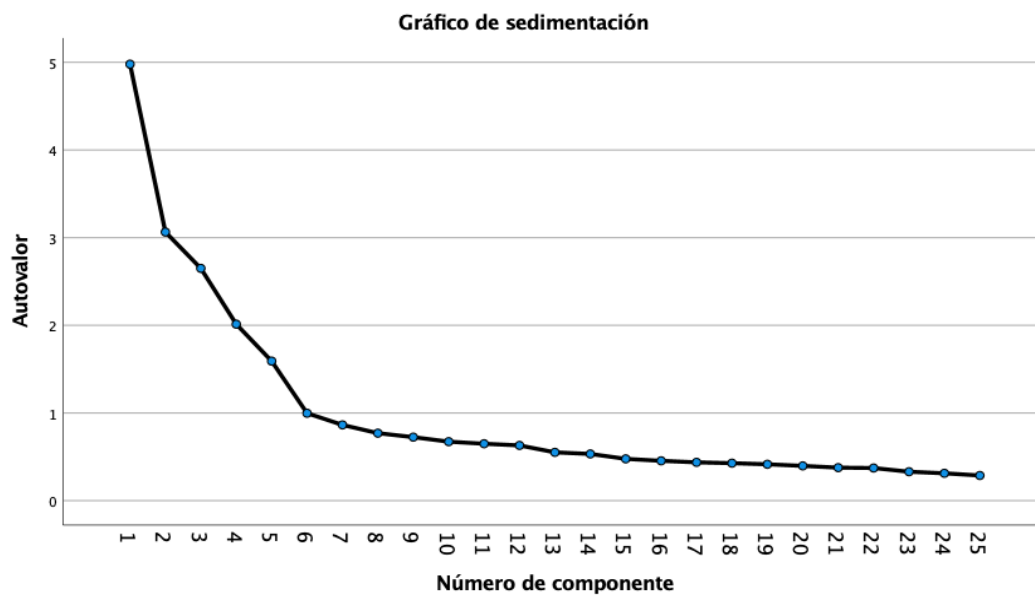
En la tabla que muestra la varianza total explicada, notamos que hay cinco autovalores mayores que 1, representando así los cinco factores que explican la mayor variabilidad.

Basándonos en la **regla de Kaiser**, opto por quedarnos con estos cinco autovalores por ser mayores que 1 y, por lo tanto, son los principales contribuyentes a la explicación de la variabilidad.

El factor más influyente es el primero, con un autovalor de 4,980, explicando el 19,918% de la variabilidad, seguido por el segundo con el 12,260%, el tercero con el 10,602%, el cuarto con el 8,057%, y el quinto con el 6,371%.

En conjunto, estos cinco factores explican el 57,208% de la variabilidad total.

Además, adjunto el gráfico de codo:



Analizando el gráfico de codo podemos justificar la elección de mínimo 5 componentes principales, ya que el comienzo de la “zona de sedimentación” comienza tras el número 5.

Por último, obtengo la matriz de componentes:

Matriz de componente					
	Componente				
	1	2	3	4	5
In your <test language lessons>, how often: The teacher poses questions that motivate students to participate actively.	0,669	0,059	0,209	-0,086	0,426
In your <test language lessons>, how often: The teacher shows students how the information in texts builds on [...]	0,654	0,068	0,228	-0,059	0,500
How often during <test language lessons>: The teacher continues teaching until the students understands.	0,630	0,015	0,355	-0,223	-0,357
In your <test language lessons>, how often: The teacher encourages students to express their opinion about a text.	0,629	0,090	0,177	-0,080	0,478
How often during <test language lessons>: The teacher helps students with their learning.	0,623	0,028	0,403	-0,224	-0,413
In your <test language lessons>, how often: The teacher helps students relate the stories they read to their lives.	0,617	0,020	0,205	-0,025	0,523
How often during <test language lessons>: The teacher gives extra help when students need it.	0,608	0,052	0,398	-0,246	-0,387
How often during <test language lessons>: The teacher shows an interest in every student's learning.	0,577	0,026	0,380	-0,230	-0,354

La matriz de componentes proporciona los pesos de cada variable en cada una de las principales componentes del modelo. La suma de los cuadrados de los valores en cada columna concuerda con el autovalor correspondiente a la componente asociada, mientras que la suma de los cuadrados de los valores en cada fila coincide con la comunalidad de cada variable.

Podría hacer una primera agrupación de variables según la componente que más información representa. Primero tenemos las 8 primeras variables que podríamos decir que hablan sobre la docencia del profesor.

Las siguientes 9 variables describen el número de algunos objetos que los alumnos tienen en sus casas. Después podemos agrupar 5 variables que nos explican la forma de estar y trabajar de los alumnos en clase. Por último, las últimas 4 variables nos muestran los diferentes textos leídos en clase durante el último mes.

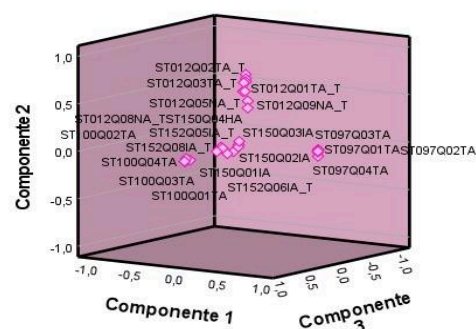
3. Análisis factorial rotado

3.1 Método varimax

La finalidad de la rotación es facilitar la interpretación de los factores. En este caso, emplearé el **método Varimax**, que es un método de rotación ortogonal. Este método implica girar los ejes de coordenadas que representan a los factores con el objetivo de minimizar el número de variables que tienen cargas altas en cada factor. En consecuencia, esto simplifica la interpretación de los factores.

3.2 Gráfico de componente en espacio rotado

Gráfico de componente en espacio rotado



En el gráfico de componente en espacio rotado observo que hay cuatro grupos, por lo que podríamos pensar que hay 4 factores, sin embargo, sabemos que hay dos conjuntos de variables relacionadas con la lectura, que son ST150 y ST152, y por eso quedan agrupadas en el gráfico, por lo que en realidad tenemos 5 factores.

3.3 Interpretación de resultados

Al realizar la *rotación varimax* obtengo la siguiente matriz de componentes rotados:

Matriz de componente rotado ^a					
	Componente				
	1	2	3	4	5
How many in your home: Televisions		,596			
How many in your home: Cars		,715			
How many in your home: Rooms with a bath or shower		,679			
How many in your home: <Cell phones> with Internet access (e.g. smartphones)		,578			
How many in your home: Computers (desktop computer, portable laptop, or notebook)		,750			
How many in your home: <Tablet computers> (e.g. <iPad>, <BlackBerry PlayBook>)		,688			
How many in your home: E-book readers (e.g. <Kindle>, <Kobo>, <Bookeen>)		,409			
How many in your home: Musical instruments (e.g. guitar, piano)		,468			
How often during <test language lessons>: Students don't listen to what the teacher says.	,787				
How often during <test language lessons>: There is noise and disorder.	,832				
How often during <test language lessons>: The teacher waits long for students to quiet down.	,831				
How often during <test language lessons>: Students cannot work well.	,774				
How often during <test language lessons>: Students don't start working for a long time after the lesson begins.	,791				
How often during <test language lessons>: The teacher shows an interest in every student's learning.			,786		
How often during <test language lessons>: The teacher gives extra help when students need it.			,838		
How often during <test language lessons>: The teacher helps students with their learning.			,856		
How often during <test language lessons>: The teacher continues teaching until the students understands.			,802		
During the last month, how often did you have to read for school: Texts that include diagrams or maps					,804
During the last month, how often did you have to read for school: Fiction (e.g., novels, short stories)					,570
During the last month, how often did you have to read for school: Texts that include tables or graphs					,833
During the last month, how often did you have to read for school: Digital texts including links					,685
In your <test language lessons>, how often: The teacher encourages students to express their opinion about a text.				,786	
In your <test language lessons>, how often: The teacher helps students relate the stories they read to their lives.				,811	
In your <test language lessons>, how often: The teacher shows students how the information in texts builds on [...]				,828	
In your <test language lessons>, how often: The teacher poses questions that motivate students to participate actively.			,239	,775	

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

a. La rotación ha convergido en 5 iteraciones.

- Parece que tenemos 5 factores que podemos clasificar como:

F1: Número de posesiones de los estudiantes.

F2: Conducta de los estudiantes.

F3: Preocupación del profesor con los estudiantes.

F4: Lectura de los estudiantes.

F5: Motivación del profesor a los alumnos.

Las variables ST012Q pertenecen al factor 2, las variables ST097Q al factor 1, las variables ST100Q0 al factor 3, las variables ST150Q0 como factor 5 y las variables ST152Q0 como factor 4.

4. Segmentación por sexo

Voy a dividir los datos por género para evaluar la validez de nuestro modelo en dos poblaciones diferentes.

La finalidad de esta segmentación es comparar las respuestas entre grupos distintos. Anteriormente, apliqué este enfoque a nivel de países. Ahora, realizo la segmentación por género, comparando las respuestas entre alumnos y alumnas. De esta manera, podré identificar las diferencias en sus opiniones.

4.1 Interpretación de resultados

Prueba de KMO y Bartlett ^a		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,864
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	2158436,952
	gl	300
	Sig.	,000
a. Student (Standardized) Gender = Female		

Prueba de KMO y Bartlett ^a		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,864
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	2170980,562
	gl	300
	Sig.	,000
a. Student (Standardized) Gender = Male		

Como podemos observar el índice KMO es muy próximo a 1 en ambos casos y el test de esfericidad de Bartlett presenta un p-valor de 0. Como ya he explicado anteriormente hay suficientes argumentos para poder hacer un análisis factorial.

Componente	Total	Autovalores iniciales	
		% de varianza	% acumulado
1	4,977	19,909	19,909
2	3,069	12,278	32,187
3	2,653	10,612	42,799
4	2,010	8,041	50,840
5	1,588	6,352	57,192
6	,999	3,996	61,188
7	,862	3,447	64,636
8	,773	3,090	67,726
9	,726	2,903	70,629
10	,675	2,701	73,330
11	,649	2,596	75,926
12	,630	2,520	78,446
13	,555	2,218	80,664
14	,536	2,143	82,807
15	,476	1,905	84,713
16	,454	1,816	86,528
17	,440	1,761	88,289
18	,428	1,710	89,999
19	,416	1,664	91,664
20	,399	1,596	93,260
21	,379	1,515	94,775
22	,375	1,500	96,275
23	,330	1,320	97,595
24	,314	1,254	98,850
25	,288	1,150	100,000

Método de extracción: análisis de componentes principales.
a. Student (Standardized) Gender = Female

Componente	Total	Autovalores iniciales	
		% de varianza	% acumulado
1	4,982	19,927	19,927
2	3,061	12,242	32,169
3	2,648	10,593	42,763
4	2,018	8,073	50,836
5	1,597	6,389	57,225
6	,998	3,992	61,217
7	,868	3,473	64,690
8	,769	3,076	67,766
9	,726	2,902	70,668
10	,673	2,690	73,359
11	,651	2,606	75,964
12	,632	2,529	78,493
13	,550	2,198	80,691
14	,532	2,128	82,819
15	,479	1,917	84,736
16	,457	1,828	86,565
17	,436	1,745	88,309
18	,429	1,718	90,027
19	,416	1,664	91,691
20	,396	1,585	93,276
21	,376	1,505	94,781
22	,372	1,486	96,267
23	,332	1,328	97,595
24	,313	1,251	98,846
25	,289	1,154	100,000

Método de extracción: análisis de componentes principales.
a. Student (Standardized) Gender = Male

Me quedo con cinco componentes principales en ambos casos por el criterio de Kaiser al ser autovalores mayores que 1. Podría pensar en quedarme con 6 por ser el sexto autovalor muy próximo a 1, pero me quedaré con 5 porque considero que ya hay suficiente variabilidad explicada.

Podemos observar que la varianza total explicada en ambos sexos es muy similar además de que cada factor aporta un porcentaje prácticamente igual a la suma total. Como en el modelo original, vemos que tiene 5 factores cada uno.

Matriz de componente rotado ^{a,b}					
	Componente				
	1	2	3	4	5
How often during <test language lessons>: There is noise and disorder.	,832				
How often during <test language lessons>: The teacher waits long for students to quiet down.	,832				
How often during <test language lessons>: Students don't start working for a long time after the lesson begins.	,793				
How often during <test language lessons>: Students don't listen to what the teacher says.	,787				
How often during <test language lessons>: Students cannot work well.	,774				
How many in your home: Computers (desktop computer, portable laptop, or notebook)		,748			
How many in your home: Cars		,717			
How many in your home: <Tablet computers> (e.g. <iPad>, <BlackBerry PlayBook>)		,687			
How many in your home: Rooms with a bath or shower		,680			
How many in your home: Televisions		,595			
How many in your home: <Cell phones> with Internet access (e.g. smartphones)		,575			
How many in your home: Musical instruments (e.g. guitar, piano)		,468			
How many in your home: E-book readers (e.g. <Kindle>, <Kobo>, <Bookeen>)		,411			
How often during <test language lessons>: The teacher helps students with their learning.			,857		
How often during <test language lessons>: The teacher gives extra help when students need it.			,838		
How often during <test language lessons>: The teacher continues teaching until the students understands.			,803		
How often during <test language lessons>: The teacher shows an interest in every student's learning.			,785		
In your <test language lessons>, how often: The teacher shows students how the information in texts builds on [...]				,828	
In your <test language lessons>, how often: The teacher helps students relate the stories they read to their lives.				,812	
In your <test language lessons>, how often: The teacher encourages students to express their opinion about a text.				,786	
In your <test language lessons>, how often: The teacher poses questions that motivate students to participate actively.			,238	,776	
During the last month, how often did you have to read for school: Texts that include tables or graphs					,834
During the last month, how often did you have to read for school: Texts that include diagrams or maps					,804
During the last month, how often did you have to read for school: Digital texts including links					,685
During the last month, how often did you have to read for school: Fiction (e.g. novels, short stories)					,571
Método de extracción: análisis de componentes principales. Método de rotación: Varimax con normalización Kaiser. ^{a,b}					
a. Student (Standardized) Gender = Male					
b. La rotación ha convergido en 5 iteraciones.					

Matriz de componente rotado ^{a,b}					
	Componente				
	1	2	3	4	5
How often during <test language lessons>: There is noise and disorder.	,832				
How often during <test language lessons>: The teacher waits long for students to quiet down.	,831				
How often during <test language lessons>: Students don't start working for a long time after the lesson begins.	,790				
How often during <test language lessons>: Students don't listen to what the teacher says.	,787				
How often during <test language lessons>: Students cannot work well.	,774				
How many in your home: Computers (desktop computer, portable laptop, or notebook)		,751			
How many in your home: Cars		,713			
How many in your home: <Tablet computers> (e.g. <iPad>, <BlackBerry PlayBook>)		,689			
How many in your home: Rooms with a bath or shower		,679			
How many in your home: Televisions		,597			
How many in your home: <Cell phones> with Internet access (e.g. smartphones)		,582			
How many in your home: Musical instruments (e.g. guitar, piano)		,468			
How many in your home: E-book readers (e.g. <Kindle>, <Kobo>, <Bookeen>)		,407			
How often during <test language lessons>: The teacher helps students with their learning.			,856		
How often during <test language lessons>: The teacher gives extra help when students need it.			,839		
How often during <test language lessons>: The teacher continues teaching until the students understands.			,801		
How often during <test language lessons>: The teacher shows an interest in every student's learning.			,787		
In your <test language lessons>, how often: The teacher shows students how the information in texts builds on [...]				,828	
In your <test language lessons>, how often: The teacher helps students relate the stories they read to their lives.				,811	
In your <test language lessons>, how often: The teacher encourages students to express their opinion about a text.				,786	
In your <test language lessons>, how often: The teacher poses questions that motivate students to participate actively.			,240	,775	
During the last month, how often did you have to read for school: Texts that include tables or graphs					,832
During the last month, how often did you have to read for school: Texts that include diagrams or maps					,803
During the last month, how often did you have to read for school: Digital texts including links					,685
During the last month, how often did you have to read for school: Fiction (e.g. novels, short stories)					,569
Método de extracción: análisis de componentes principales. Método de rotación: Varimax con normalización Kaiser. ^{a,b}					
a. Student (Standardized) Gender = Female					
b. La rotación ha convergido en 5 iteraciones.					

Como conclusión de esta segmentación para validar el modelo, podemos ver que los datos se comportan de manera muy similar para ambas poblaciones.

Apenas hay diferencia entre los valores obtenidos para mujeres y hombres, por lo que puedo concluir que no es una variable determinista.

5. Conclusiones finales

Después de haber preparado la base de datos para el análisis factorial, llego a la conclusión de que las variables pueden ser representadas por 5 factores:

- Factor 1: Comportamiento de los estudiantes en clase.
- Factor 2: Cantidad de bienes en casas de los estudiantes.
- Factor 3: El interés del profesor por los estudiantes.
- Factor 4: El profesor incentiva la lectura y su análisis en el aula.
- Factor 5: Frecuencia de lectura en el colegio.

A lo largo del análisis, he observado una correlación significativa entre las variables que pertenecen al factor 3 y 4, incluso con una variable que se encuentra en ambos factores.

Al examinar el gráfico de componentes en el espacio rotado, se observa una marcada correlación entre el factor 4 y el 5. Podría deducirse que, cuando se lleva a cabo una lectura regular de textos o libros en el entorno educativo, es común que el profesor dedique tiempo a explicarlos, motivando a los alumnos a analizarlos y compartir sus opiniones. De manera análoga, en situaciones donde no se incorpora con regularidad la lectura de textos o libros en las clases, el profesor no les concede demasiada importancia.