

Análisis de componentes principales.

El análisis de componentes principales es una técnica exploratoria de estadística multivariada que nos permite tomar un conjunto de datos (que tenga más sujetos que variables) y reducir la cantidad de variables numéricas continuas que se tienen a un número más compacto de nuevas variables que sean un concentrado de las variables originales, esto con la finalidad de poder concentrar la mayor cantidad de la variabilidad contenida en los datos en una cantidad reducida pequeña de variables que puedan ser manejadas e interpretadas con una mayor facilidad.

Las nuevas variables son combinaciones lineales de las anteriores y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra de tal modo que la componente 1 contiene mayor cantidad de variabilidad explicada que la 2, la componente 2 tiene mayor cantidad de variabilidad explicada que la 3 y así sucesivamente.

Es de vital importancia que las variables presenten cierto grado de correlación entre ellas.

Mapa de calor de Correlación



A través de la matriz de correlaciones podemos observar como existe un buen grado de asociación entre las variables tanto negativo como positivo. Llama la atención la alta correlación positiva que se presenta entre la variable X6 y X4.

El test de esfericidad de Bartlett

El test de esfericidad de Bartlett es un buen indicador respecto a que tan buenas son las correlaciones entre nuestros datos, recordemos que en principio nos interesa que exista un

buen grado de correlación entre los mismos para que al momento de aplicar esta técnica, las componentes tengan mayor peso por parte de distintas variables en cada una de ellas.

A través de esta prueba es que planteamos el siguiente juego de hipótesis:

H0: La matriz de correlacion es igual a la identidad.

Ha: La matriz de correlacion es distinta a la identidad.

Idealmente, nosotros esperaríamos encontrar valores de $p < 0.05$ para rechazar la hipótesis nula en favor de la alterna y de esta manera asegurar que encontraremos información relevante al aplicar la técnica.

Prueba de Esfericidad de Bartlett

χ^2	gl	p
567	21	< .001

A través de la prueba de Bartlett se obtuvo un estadístico $\chi^2=567$ con 21 grados de libertad y un p-valor menor a 0.05, por lo que podemos concluir que la matriz de correlaciones de nuestros datos resulta distinta a la identidad. (Esto se refiere a que, si no hubiera buenas correlaciones entre las variables, entonces la matriz de correlaciones tendería a ser una matriz identidad con sus diagonal principal con valores 1 y sus demás espacios con valores cercanos a 0)

Prueba de KMO

La prueba de Kaiser-Meyer-Olkin es una medida estadística para determinar qué tan adecuados son los datos para el análisis factorial. La prueba mide la adecuación del muestreo para cada variable en el modelo y el modelo completo.

El estadístico KMO se interpreta de 0 a 1 donde los valores más cercanos a 1 nos indican que los datos son buenos para llevar a cabo el análisis, los valores menores a 0.6 resultan malos o mediocres y a partir de 0.5 resultan ser inaceptables.

Medida de Idoneidad del Muestreo KMO

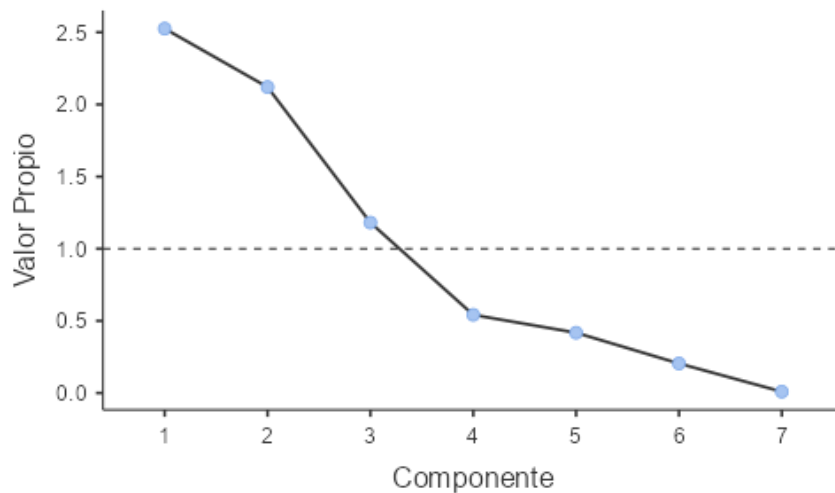
	MSA
Global	0.446
X1_VE	0.345
X2_NP	0.330
X3_FP	0.913
X4_IF	0.558
X5_SC	0.288
X6_IF	0.552
X7_CP	0.927

Tras realizar los cálculos del coeficiente KMO para la medida de adecuación del muestreo, podemos observar como existen un par de variables que presentan coeficientes de adecuación muy altos, las cuales son X3 y X7, luego están otras que logran alcanzar un coeficiente apenas aceptable como la variable X6 y la variable X4.

Finalmente vemos que hay una cantidad considerable de variables que presentan valores muy por debajo de lo mínimamente aceptable las cuales son X2,X5,X1 y Global.

Gráfica de Sedimentación

El grafico de sedimentación es una técnica visual que nos permite observar a partir de qué punto si aumentamos la cantidad de componentes principales a calcular empieza a resultar No significativo para la cantidad de variabilidad que retienen.



En el grafico se observa como por encima de la línea punteada nos indica que con 3 componentes principales seria más que suficiente si quisiéramos llevar a cabo el análisis. Si quisiéramos aumentar a 4 o más componentes el grafico nos indica que es tan poca la variabilidad que logran retener que no resultan particularmente significativas o de interés aumentar el número de componentes, con 3 resulta optimo.

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.38	34.0	34.0
2	1.83	26.1	60.1
3	1.62	23.2	83.2

El summary o resumen nos muestra la cantidad de carga que absorbe cada componente principal.

Así como la proporción de varianza que contiene y la proporción de varianza acumulada.

En la tabla podemos observar como la componente 1 logra contener un 34% de la variabilidad total de los datos, la componente 2 logra contener un 26.1% de variabilidad y la componente 3 un 23.2% de variabilidad logrando así que entre las 3 componentes expliquen un 83.2% de la variabilidad total contenida en los datos, resulta una proporción de variabilidad muy buena, aunque quizás pueda ser mejorada.

¿Qué pasa si le quitamos la variable X5 y volvemos a calcular todo?

Suele suceder que nos interesa ver lo que ocurre cuando quitamos una variable del conjunto que posiblemente no sea de nuestro interés o que nos esté generando más ruido de lo que nos estaría ayudando a explicar, por eso se opta por excluirla y ver que ocurre.

Al quitar la variable de menos peso (X5)

Prueba de Esfericidad de Bartlett

χ^2	gl	p
206	15	< .001

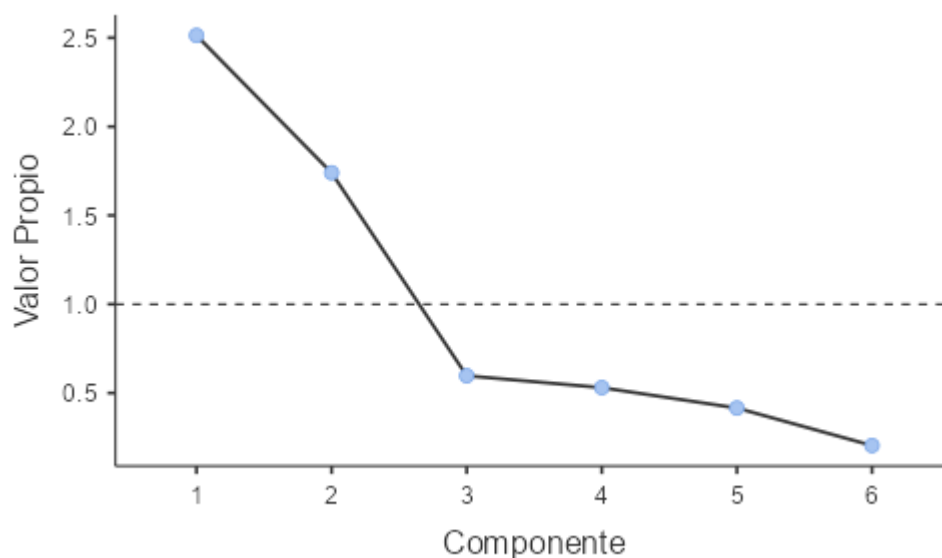
A través de la prueba de Bartlett se obtuvo un estadístico $\chi^2=206$ con 15 grados de libertad y un p-valor menor a 0.05, por lo que podemos concluir que la matriz de correlaciones de nuestros datos resulta distinta a la identidad.

Medida de Idoneidad del Muestreo KMO

MSA	
Global	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532
X7_CP	0.779

Nótese como al eliminar una variable y volver a aplicar la prueba de adecuación del muestreo con el criterio KMO ahora todos los valores resultan mayores a 0.5, indicándonos así que todos los valores se encuentran por encima del límite de lo aceptable (aunque recordemos que valores por debajo de 0.6 resultan aceptables pero mediocres)

Gráfica de Sedimentación



El grafico de sedimentación ahora nos indica que solo con 2 componentes principales podríamos explicar la mayor cantidad de la variabilidad contenida en los datos a raíz de que sacamos una variable que posiblemente nos estaba generando mucho ruido.

Resumen

Componente	SC Cargas	% de la Varianza	% Acumulado
1	2.51	41.9	41.9
2	1.74	29.0	70.9

Al obtener el resumen de las componentes principales calculadas podemos observar que el primer componente está explicando un 41.9% de la variabilidad contenida en los datos y la componente 2 está explicando un 29% de la variabilidad contenida en los datos, haciendo así que entre las 2 expliquen un acumulado de 70.9% de la variabilidad contenida en los datos.

Cargas de los componentes

Las cargas de los componentes vendrían a ser como la influencia que tiene cierta variable en cada componente., esto se puede intuir viendo los valore de las cargas que tiene una variable en cada componente y comparando el valor absoluto de la misma para saber en cual de las componentes tienen mayor peso.

Cargas de los Componentes

	Componente		Unicidad
	1	2	
X1_VE	-0.627	0.515	0.342
X2_NP	0.759		0.420
X3_FP	-0.730	0.336	0.355
X4_IF	0.494	0.798	0.119
X6_IF	0.425	0.832	0.128
X7_CP	0.766		0.385

Los resultados que se presentan son los obtenidos tras aplicar la técnica según lo planteado y sin realizar ninguna rotación.

Vemos que en las cargas de las componentes se puede notar que en la primera componente tienen mayor peso las variables X1, X2, X3 y X7 mientras que la componente 2 contiene mayor peso en las variables X4 y X5.

Introducción a las rotaciones.

Cuando no estamos muy convencidos de los resultados obtenidos por defecto tras aplicar la técnica de componentes principales entonces podemos optar por aplicarle operaciones especiales que cumplen ciertos criterios para cambiar los resultados obtenidos y verlos desde otra perspectiva sin perder alterar los datos que se están analizando.

Una analogía para entender lo que hacen las rotaciones es como si tuviéramos un cubo rubik de 6 lados y le damos la vuelta para ver que hay del otro lado del mismo., en esencia seguimos

teniendo el mismo objeto sin alterarle nada, únicamente cambiamos la perspectiva desde donde lo estamos viendo.

Rotación Varimax

Component Loadings

	Component		Uniqueness
	1	2	
X1_VE	-0.788		0.342
X2_NP	0.714		0.420
X3_FP	-0.803		0.355
X4_IF		0.933	0.119
X6_IF		0.934	0.128
X7_CP	0.764		0.385

Note. 'varimax' rotation was used

Aquí se pueden apreciar cambios significativos ya que los valores de la componente 1 aumentaron por lo cual serán mejor representadas

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.37	39.5	39.5
2	1.88	31.4	70.9

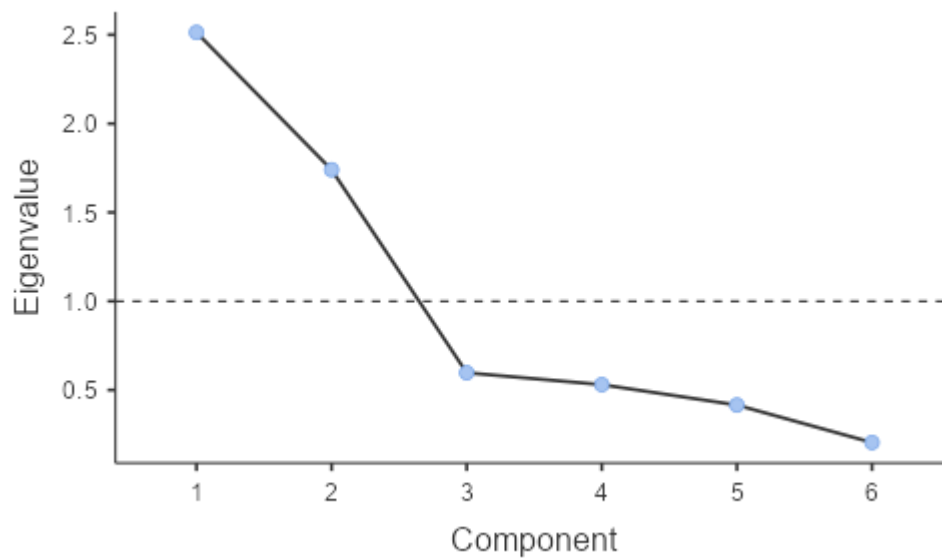
Bartlett's Test of Sphericity

χ^2	df	p
206	15	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532
X7_CP	0.779

Scree Plot



Rotación quartimax

Component Loadings

	Component		
	1	2	Uniqueness
X1_VE	0.790		0.342
X2_NP	-0.711		0.420
X3_FP	0.803		0.355
X4_IF		0.934	0.119
X6_IF		0.934	0.128

Component Loadings

	Component		Uniqueness
	1	2	
X7_CP	-0.762		0.385

Note. 'quartimax' rotation was used

En esta rotación Podemos observar que hay aún mas cambios que la anterior a que hay menos valores negativos, ademas resultan ser mas altos.

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.36	39.4	39.4
2	1.89	31.5	70.9

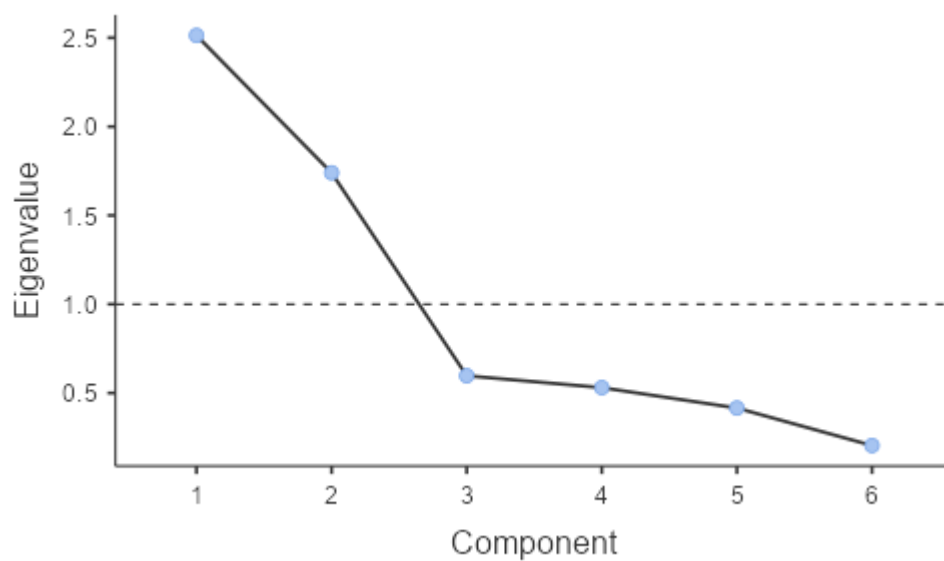
Bartlett's Test of Sphericity

χ^2	df	p
206	15	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532
X7_CP	0.779

Scree Plot



Rotación promax

Component Loadings

	Component		Uniqueness
	1	2	
X1_VE	0.813		0.342
X2_NP	-0.696		0.420
X3_FP	0.810		0.355
X4_IF		0.936	0.119
X6_IF		0.942	0.128
X7_CP	-0.754		0.385

Note. 'promax' rotation was used

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.37	39.4	39.4
2	1.89	31.4	70.9

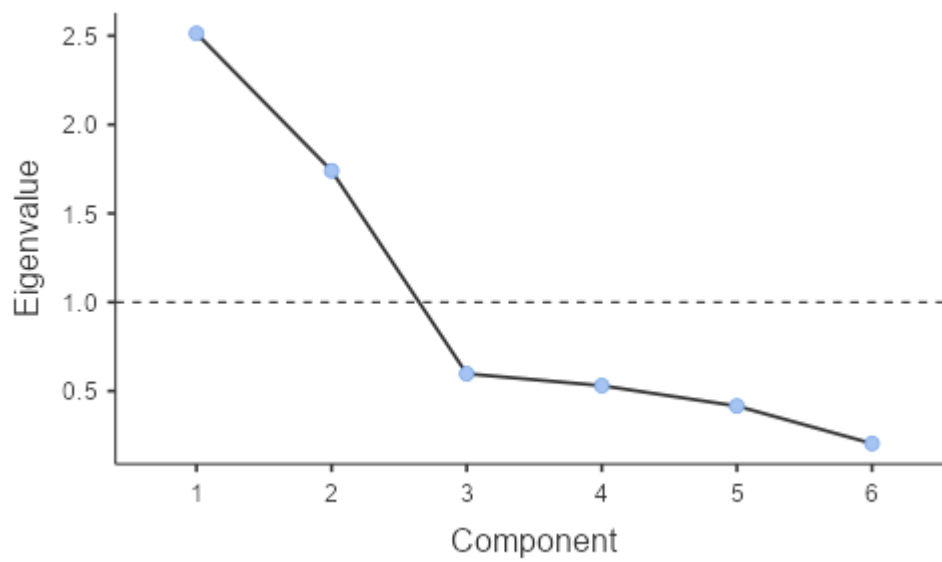
Bartlett's Test of Sphericity

χ^2	df	p
206	15	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532
X7_CP	0.779

Scree Plot



Rotación oblimin

Component Loadings

	Component		Uniqueness
	1	2	
X1_VE	0.803		0.342
X2_NP	-0.700		0.420
X3_FP	0.806		0.355
X4_IF		0.933	0.119
X6_IF		0.937	0.128
X7_CP	-0.755		0.385

Note. 'oblimin' rotation was used

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.37	39.4	39.4
2	1.89	31.4	70.9

Bartlett's Test of Sphericity

χ^2	df	p
206	15	< .001

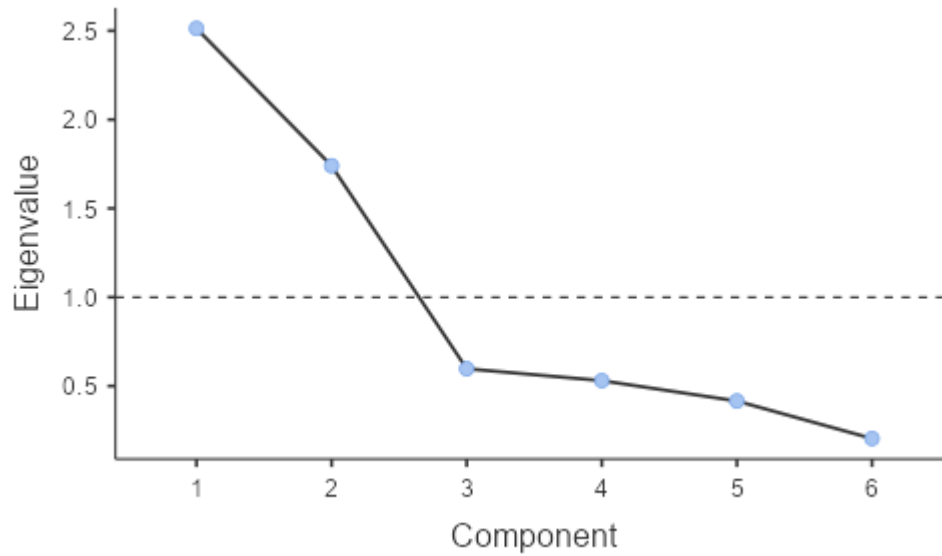
KMO Measure of Sampling Adequacy

	MSA
Overall	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532

KMO Measure of Sampling Adequacy

	MSA
X7_CP	0.779

Scree Plot



Rotación simplimax

Component Loadings

	Component		
	1	2	Uniqueness
X1_VE	0.805		0.342
X2_NP	-0.703		0.420
X3_FP	0.808		0.355
X4_IF		0.931	0.119
X6_IF		0.937	0.128
X7_CP	-0.759		0.385

Note. 'simplimax' rotation was used

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	2.38	39.6	39.6
2	1.88	31.3	70.9

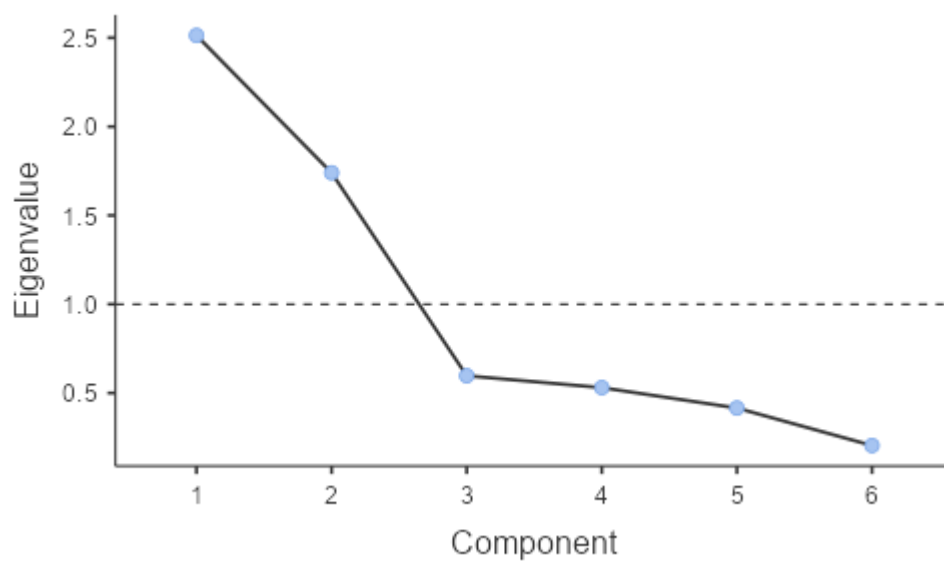
Bartlett's Test of Sphericity

χ^2	df	p
206	15	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.665
X1_VE	0.721
X2_NP	0.788
X3_FP	0.748
X4_IF	0.542
X6_IF	0.532
X7_CP	0.779

Scree Plot



En las últimas rotaciones pudimos observar que ya no había muchos cambios significativos dado que seguían permaneciendo los valores negativos y lo único que sucedía era que se intercambiaban la carga entre unos y otros, sin embargo, se observó que a partir de la segunda rotación los valores sí aumentaron.