

# A multivariate analysis to characterize homogeneous groups of countries in Africa

by

Ismael Fernando Manga Ndong Ayingono

A thesis submitted in conformity with the requirements  
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un  
i Universidad  
Internacional  
de Andalusia  
A

September 2016

# A multivariate analysis to characterize homogeneous groups of countries in Africa

Ismael Fernando Manga Ndong Ayingono

Máster en Economía, Finanzas y Computación

Supervisores: Emilio Congregado y Concepción Román  
Universidad de Huelva y Universidad Internacional de Andalucía

2016

## Abstract

In this paper we apply a set of multivariate techniques to a database of economic, social and institutional indicators of African countries, in order to obtain different groups of homogeneous countries not only in terms of relative development but also in terms of social, institutional and potential common economic interests. The analysis might be a particularly interesting issue for policy makers, and especially for those that owning appreciate natural resources has a high exposure to international trade and which bargaining power. In this sense our results might be considered as a good basis for the creation of different types of integration among African countries i.e. in a wide range of types goes from free trade areas to Economic Union– and a starting point for the diagnostic and design of general strategies of development of this continent, including the detection of potential interactions among them.

**JEL classification:** C83, N37, N47, N57 O43, O55

**Key words:** multivariate analysis, cross-section, principal components, cluster analysis, Africa, Growth and development.

## Resumen

En este trabajo, aplicamos una batería de diferentes técnicas de análisis multivariante sobre una base de datos de indicadores económicos, sociales e institucionales de los países del continente africano, para obtener grupos homogéneos no sólo en términos de su grado de desarrollo relativo sino de sus características sociales, institucionales o por la posesión de potenciales intereses económicos comunes. El análisis podría ser de particular interés para *policy makers*, y especialmente para aquellos países que poseyendo recursos naturales especialmente atractivos presentan una elevada exposición al comercio internacional y cuyo poder de negociación frente a las grandes empresas multinacionales es muy limitado. Desde esta perspectiva nuestros resultados pueden ser considerados una Buena base para la creación de diferentes formas de integración entre los países africanos –esto es cualquier forma de integración que van desde la creación de áreas de libre comercio a la creación de uniones económicas– y un punto de partida para el diagnóstico y diseño de estrategias generales de desarrollo del continente, entre las que se incluye la detección de interacciones potenciales entre ellos.

## Acknowledgments

Tengo bastante claro que no hubiera podido acabar el máster en este momento, si no hubiese contado con el cariño, el apoyo, la energía y el tiempo de muchas personas. Por tanto, quiero expresar mi reconocimiento y gratitud a todas aquellas personas que han hecho que la realización de este trabajo fin de máster sea posible.

En primer lugar, agradezco a mis tutores por tener siempre abierta la puerta a mis dudas, por sus sugerencias y confianza durante el desarrollo de este trabajo fin de máster. Igualmente, al profesor Gegúndez quien despertó en mí el interés por las técnicas que sirven de base a este estudio. Gracias por haberme dado la oportunidad de trabajar con ustedes.

Agradezco también a mi familia todo su apoyo incondicional y los ánimos que me han dado durante todo este tiempo. En especial a mi hermano Leoncio Feliciano Esono Nzé Oyana, sin ti nada hubiera sido posible. Gracias por haber confiado en mí, me lo has demostrado todo y no puedo imaginar lo que hubiera sido de mí sin tu apoyo. Has sido mi fuente de confianza y fuerza.

Por último, quiero agradecer a mis compañeros del máster, para mí ha sido un placer compartir este año de locura con todos vosotros. Con vosotros he compartido las facturas que nos presentaba la especialidad. A todos aquellos que siguen estando cerca de mí y que le regalan a mi vida algo de ellos.

# Tabla de Contenidos

1.- Introducción	p.1
2.- Metodología y datos	p.4
2.1.- El análisis multivariante como técnica de agrupación y reducción de dimensión en los datos	p.5
2.2.- El análisis multivariante como técnica para la detección de interdependencias	p.6
2.3.- Análisis de componentes principales	p.6
2.4.-Análisis factorial	p.11
2.5.- Análisis clúster	p.13
3.- Análisis y resultados	p.15
3.1.- Análisis de componentes principales	p.15
3.2.- Análisis factorial	p.26
3.3.- Análisis gráfico de los conglomerados	p.34
4.- Conclusiones y potenciales extensiones	p.40
5.- Referencias	p.43

## Lista de Tablas

Tabla 1. Matriz de correlaciones	p.15
Tabla 2. Autovalores y autovectores	p.17
Tabla 3. Autovalores y especificidades	p.19
Tabla 4. Autovalores y varianzas acumuladas	p.22
Tabla 5. Autovalores y especificidades	p.28
Tabla 6. Rotación varimax	p.30
Tabla 7. Rotación promax	p.31

## Lista de Figuras

Figura 1. Representación gráfica de la matriz de correlaciones	p.16
Figura 2. Gráfico de sedimentación	p.18
Figura 3. Gráfico de components z1 y z2	p.21
Figura 4. Mapa 1	p.21
Figura 5. Biplot 1	p.23
Figura 6. Biplot 2	p.24
Figura 7. Biplot 3	p.24
Figura 8. Biplot 4	p.25
Figura 9. Gráfico de sedimentación de los factores	p.28
Figura 10. Mapa 2	p.34
Figura 11. Mapa 3	p.35
Figura 12. Mapa 4	p.35
Figura 13. Mapa 5	p.36
Figura 14. Mapa 6	p.36
Figura 15. Dendograma	p.38
Figura 16. Mapa 7	p.39

# 1 Introducción

Indagar en la naturaleza y causa del atraso relativo de un buen número de países africanos y sobre todo en cuáles son los motores que están gobernando los llamados milagros del crecimiento en algunos de ellos y el crecimiento del continente, en general, es uno de los grandes retos al que se enfrentan analistas y *policy makers*.

Como bien es sabido, la tasa de retorno de la inversión es mayor en África que en cualquier otro país o área geográfica en vías de desarrollo, lo que provoca que tanto la búsqueda de oportunidades de negocio como la llegada de inversores extranjeros se encuentre en plena efervescencia. Por arrojar sólo un dato, a lo largo de la pasada década la tasa de crecimiento medio se encontraba alrededor del 4,9%, boom que se concentraba especialmente en los sectores de la construcción, las telecomunicaciones, la banca y el comercio, y con ritmos parecidos a los que lo hacían economías emergentes como la brasileña o la rusa.

Además y frente a lo que parece desprenderse de algunas interpretaciones, no fue este un proceso generado tan solo por la explotación de recursos naturales<sup>1</sup>, sino que también en esta ocasión, se vio acompañado de otros motores de crecimiento y de cambios estructurales que lo propiciaron e impulsaron. Baste constatar que aquellos países que no gozan de estos recursos naturales o cuya exportación no supone una parte sustancial de su PIB, también experimentaron altos ritmos de crecimiento.

Sin embargo, este escenario de “milagros” del crecimiento coexistía, como ha ocurrido tradicionalmente en este continente, con “desastres”, naturales o no, que afectaron a países singulares y con el de la persistencia de ciertos problemas: de grandes retos pendientes entre los que al menos tendríamos que incluir, al menos, la erradicación de la pobreza, la atención a la discapacidad, o incluso la alta mortalidad infantil.

---

<sup>1</sup> No podemos obviar que durante el período expansivo anterior, las economías africanas se vieron favorecidas por la positiva evolución mostrada por los precios de algunos recursos energéticos y la mayor parte de las materias primas, que dieron un impulso inicial a muchas de estas economías, que experimentaron altos ritmos de crecimiento durante la pasada década. Hay que tener en cuenta que esta explotación de recursos naturales genera casi un tercio del crecimiento de las economías africanas.

En este sentido, no podemos obviar que el entender y explicar la entre comillas desastrosa experiencia de crecimiento de este continente es uno de los grandes retos de la teoría del crecimiento, incapaz de dar una respuesta sólida sin remitirse a los shocks exógenos causados por la inestabilidad política, económica y social (Bertocchi y Canova, 2002). Aunque en la literatura empírica de crecimiento económico podemos encontrar algunas explicaciones no parece demasiado creíble el porqué ninguno de estos países haya sido capaz de aprovechar el impulso otorgado por una coyuntura favorable en los precios de las materias primas o por el descubrimiento de nuevos recursos, para dar un impulso decisivo a su economía hacia el despegue. Repasemos brevemente algunos de estos ensayos. Barro (1991) encuentra evidencia de que en los países subsaharianos existen factores idiosincráticos que obstaculizan el crecimiento, factores que Easterly and Levine (1997) identifican con la diversidad étnica, Schmidt-Hebbel (1996) con las políticas fiscales, Sachs and Warner (1997) con la geografía política, Bertocchi y Canova (2002) con el tipo de colonización y Rodrik (1998) con la política comercial. Sea cual sea la interpretación correcta convendremos que la heterogeneidad y diversidad y la falta de estrategias comunes del continente o de grupos de países del continente, impide u obstaculiza su potencial de desarrollo, de forma que el aportar elementos técnicos a las posibles alianzas y a la formación de grupos de países con intereses comunes o que puedan plantear estrategias conjuntas de desarrollo parece uno de los elementos clave para que el continente no vuelva a experimentar su enésimo fracaso en términos de desarrollo.

África es y ha sido el continente muy heterogéneo y en el que se concentran el mayor número de países menos desarrollados. De su heterogeneidad, dan buena prueba no solo sus diferentes pautas de especialización y estrategias de desarrollo sino también, la amplia varianza en niveles y ritmos de crecimiento.

Otro de los factores que caracterizan la situación y evaluación de estas economías es el de su amplia volatilidad, en el que se combinan episodios de Milagros y desastres no solo ocasionados por guerras o catástrofes naturales sino también por la escasa inestabilidad política y la emergencia de factores que vienen a indicar la ausencia de modelos estables de crecimiento o de estrategias a largo plazo de forma que los períodos de bonanza se ven interrumpidos por Fuertes cataclismos que afectan a estas economías.

Sometidos a los vaivenes de la coyuntura internacional y sin capacidad de influir sobre ella.

La mayor parte de los milagros del crecimiento obedecen más a descubrimientos inesperados que al resultado de estrategias conscientes de crecimiento o a la aplicación de políticas orientadas a aplicar las recetas del crecimiento (explicar brevemente recetas de crecimiento). Así, los llamados Milagros y desastres del crecimiento, esto es,

Quizás la alta dependencia de la disponibilidad de recursos energéticos o primarios como base de estas economías pueda ser una de las causas de esta alta volatilidad, y la presión por el control de los mismos por parte de los grupos políticos y sociales en sus territorios, a fuente de inestabilidad política y social en las mismas.

Así, y además de modificar la estrategia y orientarse hacia la aplicación de recetas que proporcionen bases más sólidas para un crecimiento sostenido que permita salir a muchas economías de los *poverty traps* en los que se hayan inmersos, muchos de ellos, situados en etapas de desarrollo iniciales, con un modelo económico basado en la exportación de recursos naturales y con una fuerte exposición a los movimientos de la coyuntura internacional han de tratar encontrar una estrategia común y caminar hacia diferentes formas de unión económica que les permitan diseñar estrategias de desarrollo comunes así como situarse en una situación de negociación que les permita evitar enfrentarse a las grandes empresas que operan en estos mercados y que, de manera efectiva controlan los precios al disponer de un poder de mercado muy superior al de estos países, y por tanto del poder de negociación en los mercados internacionales. exportador de y con fuerte dependencia de las grandes empresas multinacionales que controlan sus mercados.

En este contexto, este trabajo aspira a desarrollar un marco que nos ayude a detectar las oportunidades, alianzas y estrategias que permitan que estos países puedan alcanzar una mayor efectividad en sus experiencias de crecimiento. Para ello, y haciendo uso de técnicas de análisis multivariante tratamos de encontrar elementos comunes subyacentes a este grupo de economías tan diversas que ha de ayudar a los agentes que toman decisiones en el marco de las empresas e instituciones de estos países a diseñar estrategias empresariales y nacionales óptimas. Se trata de un análisis exploratorio, dada la escasez de antecedentes rigurosos, y preliminar, pero que puede

resultar un útil punto de partida para entender la heterogeneidad y sus causas y detectar la existencia de posibles conglomerados para el diseño de conductas cooperativas frente a terceros.

El resto del trabajo se articula en tres secciones. La segunda realiza una descripción somera de los supuestos, utilidad y procedimiento de las técnicas de análisis utilizadas en este estudio así como de los indicadores incluidos en la base de datos sobre la que se implementan las citadas técnicas. La tercera presenta los datos y resultados obtenidos a partir de los análisis propuestos, mientras que, finalmente, el último apartado se reserva al análisis y discusión de resultados, así como a la presentación de posibles extensiones naturales de este trabajo.

## 2 Metodología y datos

Gracias a las mejoras en los sistemas de obtención y captura de información, el analista dispone de una gran cantidad de información, que no solo permite acometer análisis de mayor robustez sino que provoca que buena parte de los indicadores disponibles se solapen en lo que miden. La respuesta a preguntas complejas en las que se involucran muchas variables, muchas de las cuales no dejan de ser indicadores alternativos que capturan una misma dimensión, hacen que el uso de las técnicas de análisis multivariante se conviertan en un prerrequisito para abordar cualquier investigación de corte predictivo o el contraste de cualquier hipótesis. ignoran información bastante significativa en las investigaciones.

Como bien es sabido, el análisis multivariante incluye técnicas que nos permiten, al menos: i) describir las asociaciones o relaciones entre las variables; ii) observar las influencias de unas variables sobre otras; iii) determinar grupos con características similares; iv) e identificar componentes latentes que influyen o pueden influir en las magnitudes-objetivo.

En resumen, las técnicas de análisis multivariante analizan grandes conjuntos de datos que constan de una gran cantidad de variables medidas en gran cantidad de unidades experimentales u objetos con el fin de evidenciar la información común que presentan las variables, sus interrelaciones y descubrir su estructura latente o subyacente.<sup>2</sup>

---

<sup>2</sup> En este trabajo haremos uso de algunas de las técnicas de análisis no supervisado como el análisis de componentes principales, el análisis factorial y el análisis cluster.

Como ya hemos avanzado, en este trabajo haremos uso de algunas técnicas de análisis multivariante, que son usadas con fin exploratorio, esto es, de reducción de datos, de reducción de dimensión, aunque somos conscientes de que en una siguiente fase de este trabajo el uso de técnicas de análisis inferencial nos ha de permitir avanzar en la construcción de indicadores sintéticos que nos permitan analizar la evolución de las observaciones entre grupos e intragrupos en el tiempo.

En este trabajo, nos centraremos en tres métodos multidimensionales:

Por otra parte, cabe señalar que el análisis multivariante se ocupa de dos áreas: exploratoria e inferencial. Como hemos explicado anteriormente, el objetivo de muchas técnicas multidimensionales es la simplificación. A este propósito lo llamamos reducir dimensión y son las técnicas que desarrollaremos en este trabajo. Tales métodos son exploratorios en el sentido de que son empleados básicamente para generar hipótesis en lugar de contrastarlas.

## 2.1 El análisis multivariante como técnica de agrupación y reducción de dimensión en los datos

Quizás el primer problema con el que se encuentra el analista ante cualquier análisis masivo de datos, pase por la necesidad de reducir la dimensión de los mismos. Dada una muestra de  $n$  sujetos, caracterizados por  $p$  variables aleatorias, y con vistas a analizar el comportamiento de cada unidad experimental, respecto a estas variables, puede resultarnos prioritario el sintetizar la información que estas variables contienen, en otro grupo menor de variables. Por tanto, es este un uso que podemos hacer de las técnicas del análisis multivariante, esto es, simplificar un conjunto de datos multidimensional.

Los métodos multivariantes considerados en este trabajo, son métodos paramétricos, en particular aplicamos algunas técnicas de agrupación y reducción de dimensión. Este uso es coherente con el objetivo del trabajo, esto es el de resumir un conjunto de variables en uno nuevo de menor número de variables, con la menor pérdida de información posible. Además, tratan de agrupar unidades experimentales u observaciones, de tal manera que las observaciones de cada grupo sean lo más parecidas posible respecto a las características que miden las variables consideradas y lo más diferentes posible en relación a las observaciones de otros grupos distintos.

## 2.2 El análisis multivariante como técnica para la detección de interdependencias

Si a priori se es capaz de caracterizar o determinar a un grupo de variables como dependiente y al otro como independiente, entonces el objetivo será desentrañar si el conjunto de variables independientes afecta al conjunto de dependientes. Sin embargo, nos encontramos ante problemas en los que sea imposible distinguir conceptualmente entre variables dependientes e independientes. Nos interesa, simplemente, determinar cómo y por qué están relacionadas las variables entre ellas. Los métodos estadísticos multivariantes permiten abordar esta búsqueda de relaciones causales, por lo que son denominados también como técnicas de interdependencia. En este trabajo se aplicarán algunas de estas técnicas de interdependencia con el fin de explicar no sólo la información contenida en el conjunto de variables observadas, sino también sus interrelaciones.

## 2.3 El análisis de componentes principales

Como ya hemos advertido nuestro primer objetivo no es otro que el de tratar de describir la información contenida en el conjunto de datos de las variables retenidas en algún conjunto de variables menor que el de variables originales. Para este fin, la intuición no es otra que la de descubrir si algunas de las variables incluidas en la matriz de datos puede ser considerada combinación lineal de otras, esto es, si aporta información redundante. Por tanto, si las  $p$  variables observadas están *fuertemente correlacionadas*, será posible sustituirlas por menos variables sin gran pérdida o coste de “información”.

El análisis de componentes principales (ACP) es el primer método de interdependencia que vamos a utilizar en este trabajo. Las variables no se encuentran previamente divididas en dos subconjuntos, es decir, no existen variables dependientes e independientes. Así mismo, es una técnica de reducción de dimensión, esto es, lo que busca es reducir el número de variables. En suma, se trata de expresar la mayor cantidad de información, contenida en la matriz de datos original, en el número menor de dimensiones posible. Para ello, el análisis de componentes principales construye nuevas variables obtenidas como combinaciones lineales de las variables originales, de tal forma que estas nuevas variables expliquen la máxima información o variabilidad contenida en las variables originales con el menor coste de información posible

exigiendo además estas nuevas variables estén incorreladas entre sí. Las nuevas variables obtenidas son las llamadas *componentes principales*.

La variabilidad contenida o incorporada en las componentes principales relativa a las variables originales es decreciente, esto es, la primera componente es la que mayor información abarca o contiene, a la que le sigue la segunda componente en orden de importancia por información abarcada y así sucesivamente. Por tanto, las nuevas variables son combinaciones lineales de las originales y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recoge cada nueva variable:

$$\mathbf{S} = \begin{pmatrix} \mathbf{s}_{11} & \mathbf{s}_{21} & \dots & \mathbf{s}_{1p} \\ \mathbf{s}_{21} & \mathbf{s}_{22} & \dots & \mathbf{s}_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{s}_{p1} & \mathbf{s}_{p2} & \dots & \mathbf{s}_{pp} \end{pmatrix}$$

$$\mathbf{s}_{11} > \mathbf{s}_{22} > \mathbf{s}_{33} > \dots > \mathbf{s}_{pp} = \lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$$

Siendo  $s_{jj} = \lambda_j$ , la varianza o variabilidad explicada por cada componente.

Así, y aunque el nuevo conjunto de variables tenga la misma dimensión  $p$  que la del conjunto de las variables originales, la diferencia entre los dos conjuntos de variables radica en que, las componentes principales están incorreladas y están ordenadas en función de la mayor información abarcada o varianza. Evidentemente, la efectividad del análisis depende del grado de correlación entre las variables originales, en el sentido de que, cuando las variables están muy intercorrelacionadas, entonces la mayor parte de su variabilidad puede ser explicada con muy pocas componentes principales y por tanto, la reducción de dimensión es muy efectiva. Por otra parte, cuando el grado de correlación entre las variables es menor la reducción de dimensión no resulta tan efectiva, ya que se necesitan muchas componentes principales para explicar la mayor parte de la variabilidad de las variables originales.

### 2.3.1 Obtención de las componentes principales

Dado un conjunto de datos, es decir, una matriz de datos  $\mathbf{Y}$  de orden  $n \times p$ , nuestro interés se centra en la varianza total. Por tanto, buscamos la combinación lineal de los valores de las variables de la muestra, de tal forma que:

$$z_i = \phi_{i1}y_1 + \phi_{i2}y_2 + \phi_{i3}y_3 + \cdots + \phi_{ip}y_p$$

tenga la mayor varianza muestral posible, sujeto a la restricción de que:

$$\sum_{j=1}^p \phi_j^2 = 1$$

A los coeficientes  $\phi_i$  que definen las componentes principales se les llaman *cargas o saturaciones o loadings*. Juntas, las cargas constituyen el vector de cargas o autovector de la componente principal  $z_i$ :  $\phi_i = (\phi_{i1}, \phi_{i2}, \phi_{i3}, \phi_{i4}, \phi_{ip})$ .

La matriz  $\mathbf{A}$  es ortogonal ( $\mathbf{A}'\mathbf{A} = \mathbf{I}$ ) cuyas columnas son los autovectores o vectores propios de  $\mathbf{S}$  o de  $\mathbf{R}$ :

$$\mathbf{A} = \begin{pmatrix} \phi'_1 \\ \phi'_2 \\ \vdots \\ \phi'_p \end{pmatrix}$$

donde  $\phi_i$  es el  $i$ -ésimo autovector de la matriz  $\mathbf{S}$ . Las componentes principales son las variables transformadas:  $z_i = \phi_i'y$ ,  $i = 1, \dots, p$  en  $\mathbf{z} = \mathbf{A}\mathbf{y}$ .

El número de componentes principales a retener es  $k$ . En el caso de que las variables estén altamente relacionadas, la dimensión esencial es mucho más pequeña que  $p$  ya que unos pocos de los primeros autovalores serán grandes y el porcentaje de variabilidad explicada estará próximo a 1 para un valor pequeño de  $k$ . Por otro lado, si las correlaciones entre las variables originales son todas pequeñas, la dimensión esencial  $k$  estará cercana a  $p$  y los autovalores serán muy parecidos entre sí. En esta última situación, la utilidad de las componentes principales es pequeña y no consigue una reducción de dimensión adecuada o apreciable. En el caso extremo

en que las variables estuviesen incorreladas entonces las componentes principales coinciden con las variables originales.

### 2.3.2 Cálculo de las componentes principales a partir de la matriz de correlaciones

Las componentes principales no son invariantes ante cambios de escala y, por tanto, las unidades de medida de las variables influyen sobre ellas. Si es posible, todas las variables deben estar medidas en las mismas unidades. Si las variables tienen varianzas dispares o están en unidades de medida distintas, se pueden estandarizar antes de extraer autovalores y autovectores. Esto es equivalente a determinar las componentes principales de la matriz de correlaciones  $\mathbf{R}$ , es decir, obtener las componentes principales a partir de la matriz de correlaciones.

El que una variable tenga una varianza mucho mayor que el resto de las variables, provoca que la primera componente represente en gran medida a esa variable y las otras componentes tendrán varianzas despreciables, esto es, si las varianzas difieren notablemente o si las unidades de medida no son iguales, entonces las componentes principales de  $\mathbf{S}$  estarán dominadas por las variables con las varianzas más grandes. Estas componentes principales que se han calculado en base a la matriz de covarianzas  $\mathbf{S}$ , no implican a las otras  $p - 1$  variables y, por tanto, es preferible obtener las componentes en base a la matriz de correlaciones  $\mathbf{R}$  ya que con ella se da igual importancia a todas las variables originales.

$$\mathbf{R} = \begin{pmatrix} \mathbf{1} & r_{21} & \dots & r_{1p} \\ r_{21} & \mathbf{1} & \dots & r_{2p} \\ \vdots & \vdots & \mathbf{1} & \vdots \\ r_{p1} & r_{p2} & \dots & \mathbf{1} \end{pmatrix}$$

La matriz  $\mathbf{R}$  es una matriz cuadrada y simétrica que tiene unos en la diagonal principal y fuera de ella los coeficientes de correlación entre las variables. La matriz de correlaciones muestral es análoga a la matriz de dispersión en la que se sustituyen las covarianzas por sus respectivas correlaciones muestrales. Como indicábamos al principio, el uso de esta matriz evita el que las componentes principales de  $\mathbf{R}$  sean sensibles a cambios de escala.

### 2.3.3 Representación gráfica de las componentes principales

La representación gráfica de las componentes principales es una ilustración de la rotación de los ejes, es decir, trata de mostrar una estructura simple que favorece una interpretación fácil de las componentes principales. Por tanto, como se viene diciendo a lo largo de este trabajo, el objetivo es describir la información contenida en los datos mediante algún conjunto de variables menor  $z_i$  que el de variables originales  $y_i$ . Simplemente se evalúan las dos primeras componentes principales  $(z_1, z_2)$  para cada observación y se representan esos  $n$  puntos. La representación gráfica de las componentes principales es equivalente a una proyección de la nube de puntos  $n$  en el espacio de dimensión  $p$  sobre un plano que muestra la mayor dispersión posible de los puntos.

El gráfico de puntuaciones de componentes principales puede desenmarañar características importantes de los datos. Uno de los objetivos es detectar los datos anómalos (*outliers*) o no linealidad. En general, como se ha aclarado en las dos últimas sesiones, las primeras componentes principales son bastante sensibles a observaciones anómalas que inflan las varianzas y distorsionan las covarianzas, y las últimas son sensibles a datos atípicos que introducen dimensiones artificiales u ocultan singularidades.

Un perfeccionamiento o corrección del gráfico anterior se consigue con un *biplot*. El gráfico *biplot* es una representación simultánea de los individuos mediante puntos y las variables mediante flechas, en un sistema de coordenadas bidimensional construido en base a las dos primeras componentes principales. Permite interpretar el significado de las componentes (la primera en el eje vertical y la segunda en el eje horizontal) en base a las direcciones de las flechas. En el mismo gráfico, se valoran como parecidos los individuos cuyos puntos están próximos. De igual manera tendrán correlación positiva las variables con flechas semejantes. Asimismo, los objetos o unidades experimentales que se encuentran en la dirección de cierta flecha tendrán observaciones altas en la variable representada por la flecha. Esto es, las distancias euclídeas entre los puntos del plano en el *biplot* equivalen a las distancias de Mahalanobis entre las observaciones originales y los ángulos entre los vectores representan aproximadamente la correlación entre las variables. Así un pequeño ángulo entre dos vectores indica que las dos variables están altamente correladas (siendo el ángulo de  $0^\circ$  representativo de una correlación igual a 1), dos variables cuyos vectores formen un ángulo de  $90^\circ$  están

incorreladas y un ángulo superior a  $90^\circ$  indican que las variables están opuestamente correladas (un ángulo de  $180^\circ$  representa una correlación igual a -1).

## 2.4 El análisis factorial

El análisis factorial es un método estadístico multivariante cuyo propósito principal es desenmarañar la estructura subyacente en una matriz de datos. Para ello analiza la estructura de las interrelaciones entre un gran número de variables, no exigiendo ninguna distinción entre variables dependientes e independientes, y calcula un conjunto de dimensiones latentes, conocidas como factores, que tratan de explicar dichas interrelaciones. En este análisis se supone que las correlaciones entre variables no son aleatorias y plantea que estas interrelaciones pueden expresarse a partir de una menor dimensión de variables no observables (factores) y que éstas variables son ortogonales entre sí. El Análisis Factorial se utiliza tanto para estudiar y detectar la estructura en las relaciones entre un grupo de variables como para reducir el número de variables. Es, por lo tanto, una técnica de reducción de dimensión dado que si se cumplen sus hipótesis, la información contenida en la matriz de datos puede expresarse, sin mucho coste, en un número menor de dimensiones representadas por dichos factores. En el análisis se representan a las variables originales  $y_1, y_2, y_3, \dots, y_p$  como combinaciones lineales de unas pocas variables  $f_1, f_2, f_3, \dots, f_m$  ( $m < p$ ) llamados factores. Idealmente,  $m$  debería ser mucho más pequeño que  $p$ ; en caso contrario no habremos conseguido una reducción de dimensión efectiva. Los factores son construcciones subyacentes o variables latentes que generan a  $y_1, y_2, y_3, \dots, y_p$ , es decir, las variables originales juegan el papel de variables dependientes que se explican por variables no latentes, factores. El modelo factorial expresa cada variable como una combinación lineal de factores comunes subyacentes  $f_1, f_2, f_3, \dots, f_m$  más un término de error  $\varepsilon_i$  que recoge la parte de la variable que no es común con el resto de variables:

$$\begin{aligned} y_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + \varepsilon_1 \\ y_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + \varepsilon_2 \\ y_3 &= \lambda_{31}f_1 + \lambda_{32}f_2 + \dots + \lambda_{3m}f_m + \varepsilon_3 \\ &\vdots \\ y_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + \varepsilon_p \end{aligned}$$

El mismo modelo puede ser expresado de forma matricial:

$$\mathbf{y} = \mathbf{\Lambda f} + \boldsymbol{\epsilon}$$

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \vdots & \vdots & \dots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pm} \end{pmatrix}$$

siendo  $\Lambda$  la matriz de cargas factoriales.

Podemos considerar a los  $f_j$  como variables aleatorias que generan a las  $y_i$ . A los coeficientes  $\lambda_{ij}$  se les denomina *cargas o saturaciones factoriales o loadings* y actúan como ponderaciones, mostrando cómo cada  $y_i$  depende individualmente de los factores  $f_j$ , esto es, la carga factorial representa la correlación o la importancia de la variable para el factor correspondiente.

La varianza de cada variable se expresa de la siguiente manera:

$$\text{var}(y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2 + \psi_i = h_i^2 + \psi_i$$

donde  $h_i^2$  es la comunalidad o varianza común, esto es, la parte de la variabilidad de la variable explicada por los factores comunes y  $\psi_i$  la especificidad o la varianza única. Como ya advertimos anteriormente, este tipo de análisis admite tanto un uso exploratorio como confirmatorio. El análisis exploratorio se caracteriza porque no hay conocimiento previo sobre el número de factores y es en la aplicación donde se determina este número. En el análisis confirmatorio los factores están especificados de antemano, utilizando contrastes de hipótesis para su corroboración.

### 2.4.1 Rotación de los factores

La matriz de cargas factoriales  $\Lambda$  juega un papel importante a la hora de interpretar los factores y, si éstos son ortogonales, cuantifican el nivel y tipo de la relación entre éstos y las variables originales. Sin embargo, la matriz de cargas factoriales inicial  $\Lambda$  muestra una estructura complejamente interpretable ya que en muchas ocasiones los factores están correlacionados con casi todas las variables, lo cual provoca un cierto grado de ambigüedad. Para resolver este problema están los procedimientos de rotación factorial que, a partir de la solución inicial, buscan una aproximación al **principio de estructura simple** según el cual la matriz  $\Lambda$  debe

reunir las siguientes características: i) Cada factor debe tener unos pocos pesos  $\lambda_{ij}$  altos y los otros próximos a cero; ii) Cada variable no debe estar saturada más que en un factor; y, iii) No deben existir factores con la misma distribución, esto es, dos factores distintos deben presentar distribuciones diferentes de pesos o cargas altas y bajas.

La rotación consiste en hacer girar los ejes de coordenadas, que representan a los factores, hasta conseguir que se aproxime al máximo a las variables en que están saturados, es decir, busca una estructura simple. El número de factores en los que una variable tiene cargas moderadas o altas es denominado *complejidad* de la variable. En la situación ideal de una estructura simple todas las variables tienen complejidad 1. En este caso, las variables han sido expresamente conglomeradas en grupos correspondiente a factores.

De esta manera, y dado que hay más variables que factores, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de variables. Analizando las características de las variables de un grupo asociado a un determinado factor se pueden encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a estas características comunes.

## 2.5 Análisis clúster

El análisis clúster es también una técnica de reducción de dimensión de la información, al igual que lo eran el análisis factorial o el de componentes principales. Sin embargo, y frente a los dos anteriores, el análisis clúster no se centra en la reducción de número de variables (o columnas) en la matriz de datos, sino en reducir el número de filas es decir las unidades de observación, a través de su agrupación las observaciones en grupos homogéneos en función de las similitudes entre ellos. A estos grupos o conglomeraciones se les denomina *clústers* y a la técnica *análisis clúster o de conglomerados*.

El objetivo de este análisis es la búsqueda de patrones que en los datos que permitan aglomerar los ítems en grupos o clúster. Este agrupamiento debe ser óptimo en el sentido de que las observaciones dentro de cada clúster sean similares, pero los grupos deben ser distintos entre sí. En general el análisis clúster se emplea para agrupar las observaciones, aunque también se puede aplicar para aglomerar las variables, en cuyo caso la semejanza podría ser una correlación. En

definitiva, se trata de identificar los ítems u observaciones vectoriales que son similares y agruparlos en grupos o clústers.

## 2.6 Descripción de los datos

En coherencia con los objetivos del trabajo, la aplicación empírica se realiza sobre la recopilación de una batería de 31 indicadores económicos y sociales de una muestra de 47 países africanos para el año 2013, obtenidas del Banco Mundial. Las variables retenidas en el análisis son de estudio se describen a continuación: 1) Renta per cápita (GDP); ii) índice de desarrollo humano (Hum); iii) Inversión extranjera expresada en porcentaje del PIB; iv) Tasa de desempleo (Desem); v) rentas de trabajo (Ingr); vi) % área de bosques (Bosque); vii) índice de corrupción (Corr); viii) índice de estabilidad política (Polit); ix) Gasto en salud per cápita (salud); x) infectados por VIH (VIH); x) afectados por tuberculosis (tub); xi) población (pob); xii) Tasa de mortalidad (mortl); xiii) gasto per cápita en salud; xiv) coste de iniciar un negocio (costbusns); xv) % tierra agrícola (agriperc); xvi) libertad de inversión (libinvers); xvii) acceso al agua potable (accesoagua); xviii) reservas de petróleo (reserpetrol); xix) producción de petróleo (oilprod); xx) efectividad del gobierno (gob); xxi) reservas de divisas (resrv); xxii) esperanza de vida (Espvda); xxiii) deuda externa (deuda); xxiv) imperio de la ley (ley); xxv) calidad regulatoria (regul); xxvi) valor añadido de la industria (vlorindust); xxvii) valor añadido de la fabricación (vlorfabric); xxviii) servicios de valor añadido (servañd); ixix) valor añadido de la agricultura (vloragraric); xxx) índice de innovación (indinnov); xxxi) pasajeros de aerolíneas (aerolin). Cabe aclarar que de cara al análisis estadístico computacional, todas las variables han sido estandarizadas para evitar de esta manera los problemas de cambio de escala. El tratamiento de los datos ha sido realizado en Stata y en lenguaje de programación Matlab.

## 3 Análisis y resultados

Dedicamos este apartado a la presentación de los resultados obtenidos de la aplicación de la estrategia metodológica presentada en el apartado anterior sobre el análisis de las macromagnitudes e indicadores de nuestra muestra de 47 países africanos.

### 3.1 Análisis de componentes principales

En primer lugar, necesitamos que las variables estén correlacionadas entre sí, ya que de esta manera, la mayor parte de su variabilidad se puede explicar con muy pocas componentes y

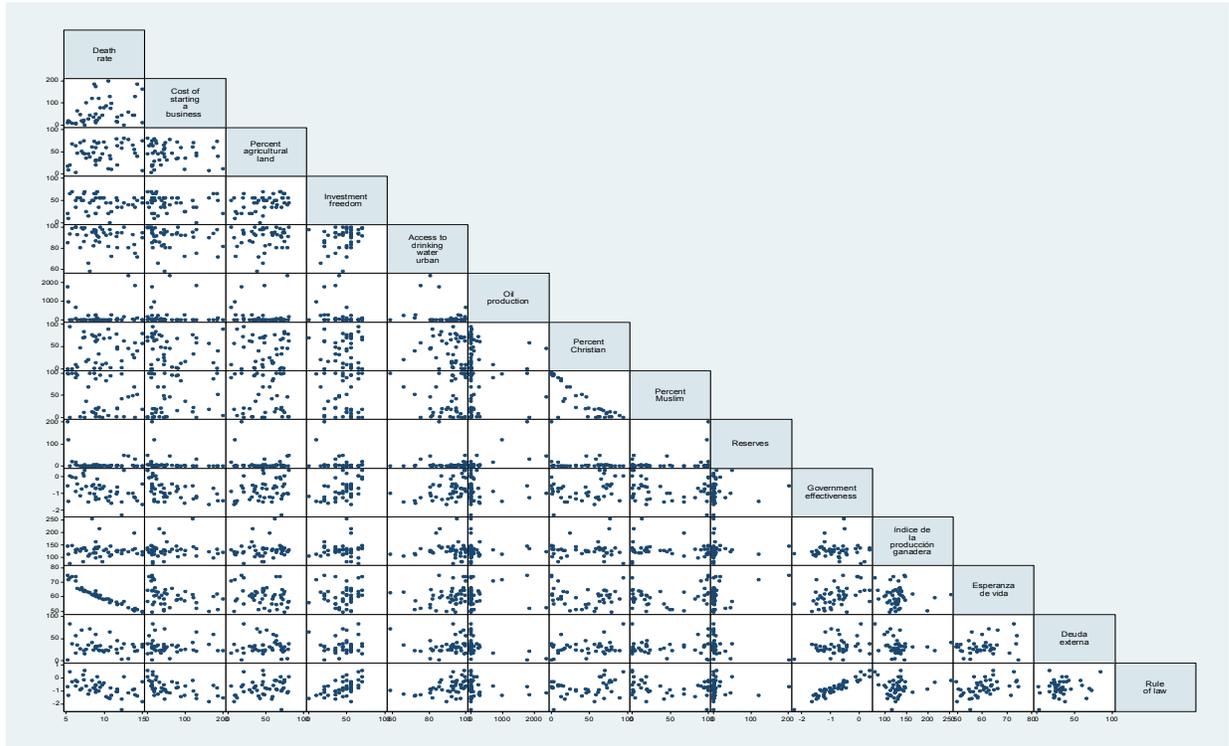
la reducción de dimensión que consigue el ACP es muy efectiva. Para ello, vamos a calcular la matriz de correlaciones y un gráfico que nos muestre las correlaciones entre todas las variables. Además de la matriz de correlaciones, en la *tabla 1*, podemos apreciar las bajas especificidades de las  $p$  variables, lo que significa que las pocas componentes retenidas explicarían muy bien la variabilidad total de los datos con una pérdida mínima de información.

Podemos observar que muchas de las variables tienen correlaciones significativas, esto es, existen patrones lineales en el conjunto  $p$  de variables. Por tanto, cabe esperar que el análisis de componentes principales resulte efectivo. Por ejemplo, al existir una relación lineal directa entre las variables control de corrupción, desarrollo humano, imperio de la ley, efectividad del gobierno, esperanza de vida y todas éstas en relación negativa con el coste de iniciar un negocio. Una notable relación inversa entre el grupo de variables renta per cápita, producción de petróleo, reservas de petróleo y la deuda de los países, lo que significaría que los países africanos productores de petróleo poseen una cierta capacidad para afrontar el pago de su deuda externa. Otra notable relación negativa entre la esperanza de vida y la tasa de mortalidad. Por tanto, de la observación de estos patrones, cabe esperar que el análisis de componentes principales sea robusto.

**Tabla 1.** Matriz de correlaciones.

	Corrp	polit	salud	ReserP-1	OilProd	Gob	ley	regul	aerolin	vlorin-t	vlorfa-c	servaãd	indinnov
Corrp	1.0000												
polit	0.6906	1.0000											
salud	0.2766	0.2726	1.0000										
ReserPetrol	-0.3557	-0.4142	0.3562	1.0000									
OilProd	-0.2888	-0.3337	0.2871	0.7377	1.0000								
Gob	0.8854	0.6616	0.3849	-0.3286	-0.2540	1.0000							
ley	0.9130	0.7156	0.3397	-0.3360	-0.3012	0.9326	1.0000						
regul	0.7900	0.6182	0.1737	-0.4007	-0.3089	0.8539	0.8548	1.0000					
aerolin	0.0958	-0.1556	0.4876	0.1308	0.2317	0.3175	0.1770	0.1729	1.0000				
vlorindust	-0.0182	-0.2859	0.3414	0.3907	0.6497	0.1104	0.0200	0.0565	0.7186	1.0000			
vlorfabric	0.0172	-0.2572	0.3620	0.3159	0.6402	0.1323	0.0286	-0.0093	0.6429	0.9131	1.0000		
servaãd	-0.0219	-0.2546	0.3193	0.3943	0.5630	0.1577	0.0410	0.1505	0.7271	0.9348	0.7612	1.0000	
indinnov	0.5669	0.4155	0.2359	-0.2272	-0.1060	0.7533	0.7019	0.6926	0.3703	0.2218	0.1781	0.2863	1.0000

**Figura 1.** Representación gráfica de la matriz de correlaciones.



Pasemos a calcular los autovalores y a presentar la varianza acumulada por cada componente. En primer lugar, se obtienen los valores propios de la matriz de correlaciones, que coincidirán con las varianzas de las diferentes componentes principales. Dichos valores vienen dados en la *tabla 2*, donde también se muestra el porcentaje de la variabilidad explicada por cada componente principal y la varianza acumulada.

**Tabla 2.** Autovalores y autovectores.

Componente	Autovalor	Diferencia	Proporción	Acumulada
Comp1	8.47662	1.05652	0.2493	0.2493
Comp2	7.4201	3.39294	0.2182	0.4676
Comp3	4.02716	.93375	0.1184	0.5860
Comp4	3.09341	1.0032	0.0910	0.6770
Comp5	2.09021	.444989	0.0615	0.7385
Comp6	1.64522	.450776	0.0484	0.7868
Comp7	1.19444	.234679	0.0351	0.8220
Comp8	.959763	.0970559	0.0282	0.8502
Comp9	.862708	.0474791	0.0254	0.8756
Comp10	.815228	.134819	0.0240	0.8996
Comp11	.68041	.0599192	0.0200	0.9196
Comp12	.620491	.144038	0.0182	0.9378
Comp13	.476453	.0843389	0.0140	0.9518
Comp14	.392114	.0965501	0.0115	0.9634
Comp15	.295564	.0280162	0.0087	0.9721
Comp16	.267548	.103558	0.0079	0.9799
Comp17	.16399	.041548	0.0048	0.9847
Comp18	.122442	.0272799	0.0036	0.9883
Comp19	.095162	.0177569	0.0028	0.9911
Comp20	.0774051	.0122809	0.0023	0.9934
Comp21	.0651243	.0197112	0.0019	0.9953
Comp22	.0454131	.0044601	0.0013	0.9967
Comp23	.040953	.00435104	0.0012	0.9979
Comp24	.0366019	.021235	0.0011	0.9990
Comp25	.015367	.0075525	0.0005	0.9994
Comp26	.00781446	.00250631	0.0002	0.9996
Comp27	.00530814	.00144072	0.0002	0.9998
Comp28	.00386742	.00144297	0.0001	0.9999
Comp29	.00242445	.00203687	0.0001	1.0000
Comp30	.000387582	.0000753646	0.0000	1.0000
Comp31	.000312217	.000312217	0.0000	1.0000

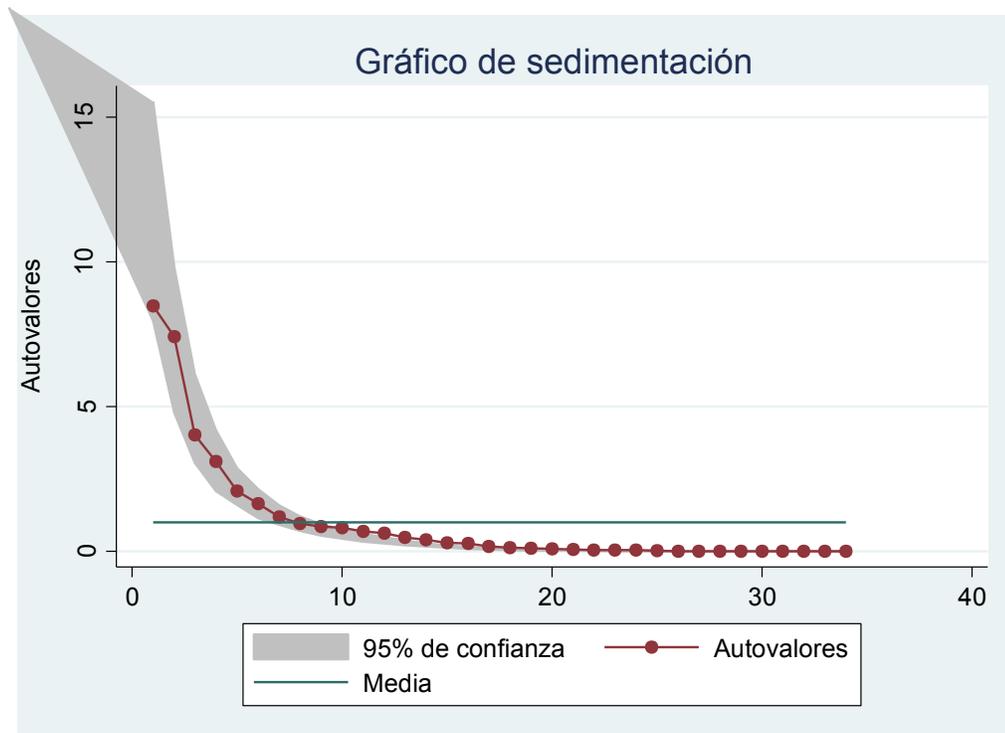
La primera componente principal logra explicar un 24.93% de la variabilidad total, la segunda componente un 21.82% y la tercera un 11.84%. Las tres conjuntamente logran explicar casi el 60% de la variabilidad total, es decir, las tres primeras componentes principales explican el 58.60% de la información contenida en las treinta y seis variables originales

La regla de *Kaiser* nos aconseja tomar el mismo número de componentes que autovalores mayores que 1, que en este caso serían siete y explicarían el 82.20% de la varianza total, aunque estudios recientes aconsejan tomar las componentes cuyos valores propios son superiores a 0'7,

por lo que podríamos retener hasta diez componentes que explicarían casi el 90 % de la variabilidad total de los datos.

Según el gráfico de sedimentación debemos retener al igual que en el criterio de *Kaiser* siete componentes principales. En la figura observamos que a partir del punto 8 desde arriba, la línea roja empieza a ser recta, figura 2. Esto significa que, además de las primeras siete componentes, el resto no contiene información significativa como para ayudar a explicar la variabilidad contenida en las 36 variables originales.

**Figura 2.** Gráfico de sedimentación



Resulta práctica común en este tipo de literatura el retener componentes que expliquen al menos el 80% de la variabilidad total. En este trabajo, vamos a considerar como óptimo *siete* componentes principales dado que explican conjuntamente el 81.88%, porcentaje por encima del mínimo recomendado.

**Tabla 3.** Autovectores y especificidades

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Unexplained
GDPper	0.2532	-0.0149	0.2230	-0.0764	0.1050	-0.1138	-0.2910	.09095
HumDII	0.2827	-0.0101	0.0631	-0.2061	0.0909	-0.0499	-0.0382	.1512
Exprt	-0.0035	-0.0826	0.3442	-0.0574	0.2413	0.1664	0.1089	.2806
Inver	-0.0609	-0.0848	0.2436	0.0548	-0.2321	0.3805	0.1964	.2701
Desem	0.1344	-0.0157	0.3320	-0.0434	-0.3876	0.0477	-0.0088	.07751
Ingr	-0.1010	0.0727	0.3445	-0.1731	0.1341	0.1686	-0.1821	.1796
Bosque	-0.0669	-0.1095	0.2408	0.0581	0.3128	0.0406	-0.2841	.3256
Corrp	0.2242	-0.2002	-0.1149	-0.0335	0.0179	0.1438	-0.0955	.1742
polit	0.1379	-0.2732	0.0786	0.0807	0.1471	0.1129	-0.0814	.166
salud	0.2859	0.0115	0.1885	-0.0460	0.0557	-0.1997	-0.0645	.0795
VIH	0.1179	-0.0866	0.2048	0.3087	0.0806	-0.1601	0.1636	.2752
tblrlsis	0.1112	-0.0006	0.3196	0.3179	0.0236	-0.1058	-0.0119	.1514
pop	0.0550	0.3014	-0.0934	0.1428	-0.0363	0.0167	-0.0561	.1949
mortl	-0.1183	0.1022	0.1403	0.4087	0.0057	-0.1097	-0.1406	.1643
costbusns	-0.2453	0.0414	-0.0039	0.0895	0.0925	-0.1400	0.0408	.4004
AgriPerc	0.0502	-0.0321	-0.1230	0.3419	-0.1004	0.1797	0.2548	.3967
libInversn	0.1209	-0.1900	-0.1596	0.1301	-0.0417	0.3428	-0.2284	.194
Agua	0.1340	-0.1093	-0.0997	0.0604	0.3575	-0.3001	0.0144	.2921
ReserPetrol	0.0707	0.2997	-0.0427	0.0806	0.0508	0.2667	-0.1911	.09794
OilProd	0.1112	0.3087	-0.0069	-0.0461	0.1113	0.2256	-0.0961	.06084
Gob	0.2731	-0.1664	-0.1004	0.0762	0.0055	0.0423	0.0316	.09957
reserv	0.1552	0.1901	0.0294	-0.1959	0.1545	0.0647	0.1979	.302
EspVda	0.1697	-0.0651	-0.1120	-0.4315	-0.0211	0.0465	0.1000	.08156
Gndria	0.0248	-0.0289	0.1176	0.0798	0.3082	0.2729	0.5816	.1881
Deuda	-0.0301	-0.1804	0.0510	-0.0727	-0.3619	-0.1139	-0.0161	.4286
ley	0.2510	-0.2033	-0.1159	0.0270	0.0442	0.1251	-0.0137	.07298
regul	0.2204	-0.1743	-0.1150	0.1885	-0.0723	0.1734	-0.1493	.1128
aerolin	0.2292	0.1326	0.0077	0.0398	-0.1463	-0.2855	0.2243	.1802
vlorindust	0.2039	0.2817	0.0140	0.0227	0.0031	0.0033	0.0231	.05556
vlorfabric	0.1981	0.2591	0.0084	-0.1078	0.0579	-0.0031	0.1548	.09746
servaRd	0.1951	0.2640	0.0061	0.1758	-0.0682	-0.0165	-0.0211	.05395
desmpjuv	0.1929	0.0065	0.2903	-0.0539	-0.3429	-0.0247	0.0382	.08745
vloragric	0.0761	0.3064	-0.1025	0.1136	-0.0235	0.1839	-0.1541	.08695
indinnov	0.2400	-0.0887	-0.1999	0.1544	0.0301	-0.1352	0.0588	.1825

Los vectores propios asociados a las dos primeras componentes principales son:

$$\phi_1 = (0.25, 0.28, \dots, 0.2242, \dots, 0.2859, \dots, -0.2453, \dots, 0.2731, \dots, 0.2510, \\ \dots, 0.2204, 0.2292, \dots, 0.24)$$

$$\phi_2 = (\dots, -0.2732, \dots, 0.3014, \dots, 0.30, -0.1900, 0.3087, 0.20, -0.1804, \\ \dots, -0.2033, 0.20, \dots, -0.1804, \dots, 0.2817, \dots, 0.26)$$

por lo que estas dos componentes son:

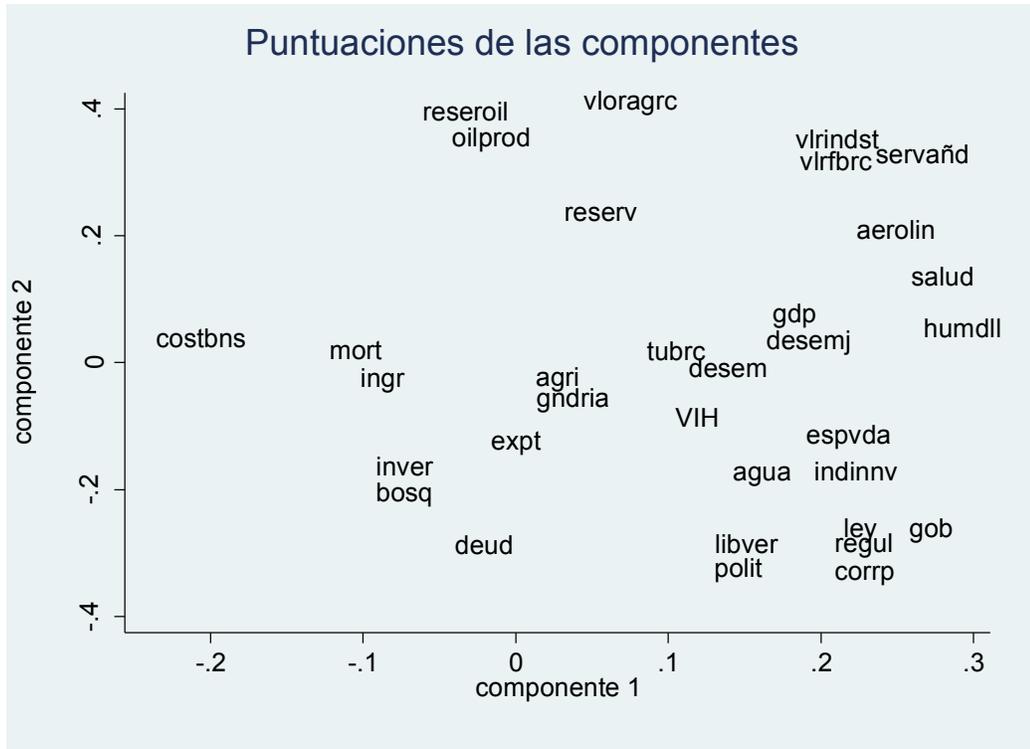
$$z_1 = 0.25GDPper + 0.28HumDII + 0.2242Corrp + 0.286salud \\ - 0.2453costbusns + 0.2731Gob + 0.2510ley + 0.2204regul \\ + 0.23aerolin + 0.2400indinnov$$

$$z_2 = -0.2732polit + 0.3014pop + 0.30ReserPetrol - 0.1900libInversn$$

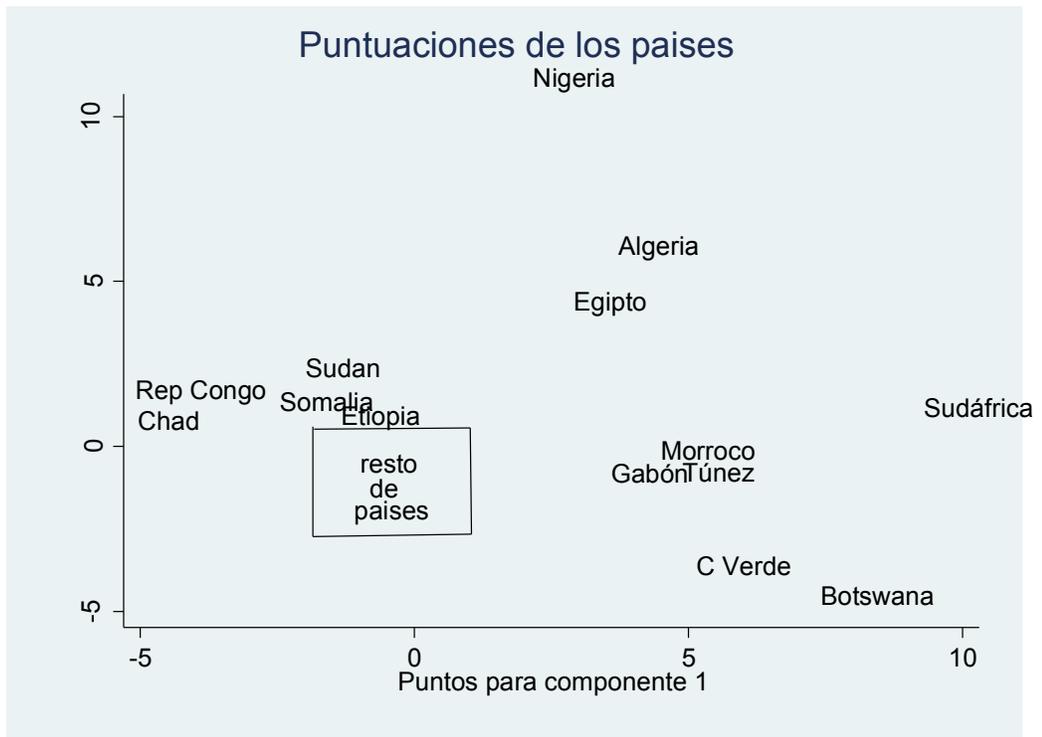
$$\begin{aligned}
 &+ \mathbf{0.3087 OilProd} + \mathbf{0.1901 reserv} - \mathbf{0.1804 Deuda} - \mathbf{0.2033ley} \\
 &+ \mathbf{0.2817 vlorindust} + \mathbf{0.3064 vloragric}.
 \end{aligned}$$

Observamos que los coeficientes más grandes, esto es, las cargas más grandes en  $z_1$  y  $z_2$ , resultados en negrita, corresponden a las variables que tienen mayor variabilidad. Es decir, estas variables tienen mucha influencia sobre las dos componentes, esto es, son las que más ayudan y aportan a la capacidad explicativa de  $z_1$  y  $z_2$ . La *primera componente*  $z_1$  explicaría las características relacionadas al desarrollo económico de algunos países del continente y, la *segunda componente*  $z_2$  describe las características relacionadas con la riqueza natural y el valor añadido. En la *figura 3*, se muestra gráficamente la capacidad explicativa de las dos primeras componentes principales. En la *figura*, vemos por ejemplo, que la segunda componente sitúa en la parte superior las características como las reservas de los países, reservas de petróleo, producción de petróleo e ingresos de los recursos naturales y, sitúa en la parte baja del gráfico la deuda externa, lo que mostraría la capacidad de alivio de deudas de los países africanos productores de petróleo. La primera componente principal sitúa en la parte derecha las características como relacionadas con el marco jurídico, la gobernabilidad y la estabilidad política y, en la parte izquierda ilustra las características como la mortalidad y el coste de iniciar un negocio, por lo que, esta componente describe el desarrollo económico de algunos países africanos. Al igual que la figura anterior, *el mapa 1* hace agrupaciones significativas de países.

**Figura 3.** Gráfico de componentes  $z_1$  y  $z_2$



**Figura 4.** Mapa 1

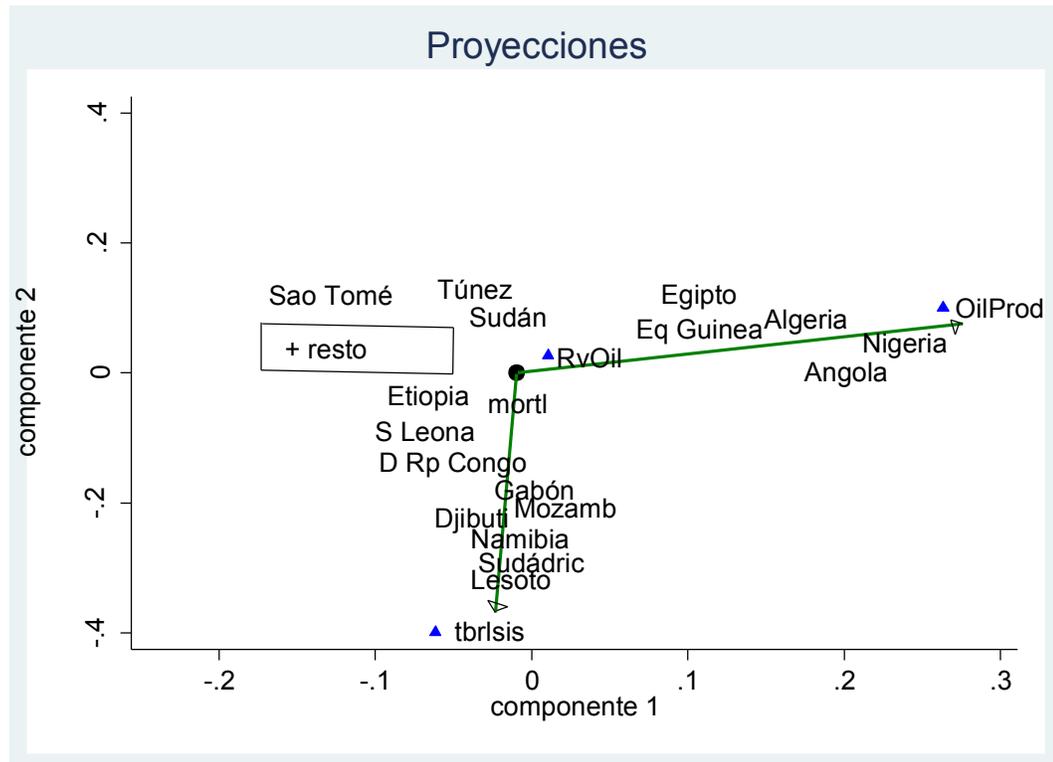


En la *figura 4*, hemos utilizado las dos primeras componentes principales para hacer una proyección de las similitudes entre algunos países.

Diversas características son detectadas: las observaciones correspondientes a Nigeria y Argelia son unos outliers ya que aparecen desplazadas en relación al resto, esto es debido a sus altos valores en la producción y reservas de petróleo, como todos sabemos, Nigeria es el mayor productor de petróleo de África y Argelia es otro muy importante. En el centro de los ejes está la mayor concentración de los países, lo que nos hace pensar que los países africanos a pesar de poseer características importantes que les harían diferentes de los unos a los otros, pero sin embargo, las magnitudes sociales, políticas e institucionales hacen de ellos países muy similares. Este grupo lo formarían los países que se destacan menos en las características relacionadas a la riqueza natural y están muy caracterizados por el malestar social como es la tasa de mortalidad, el costo de iniciar un negocio, alta dependencia del sector agrícola, alta deuda externa, en definitiva, este grupo lo forman gran parte de los países del continente, el grupo de los países más pobres del continente. En la parte derecha tenemos otro grupo en el que se destacan las características como esperanza de vida, control de corrupción, control político, efectividad de gobierno, desarrollo humano, acceso al agua potable, libertad de inversión, esto es, países en los que se destaca el desarrollo económico. Los siguientes *biplots* nos aclaran mejor este fenómeno.

En la *figura 5*, los países localizados a la derecha del gráfico son aquellos en los que predomina el sector petrolero, esto es, las reservas y/o producción de petróleo como es Nigeria, Angola, Argelia, Guinea Ecuatorial, República de Congo, Egipto etc. Los países situados en la parte inferior del gráfico como son Sudáfrica, Djibouti, Namibia, Lesoto y Mozambique son países con fuerte infección de tuberculosis y altos índices de mortalidad.

Figura 5. Biplot 1



Algunas de las incongruencias del continente son como la que podemos observar en la *figura 6*, donde Guinea Ecuatorial se destaca por tener la mayor renta per cápita del continente y Nigeria el mayor productor de petróleo, pero sin embargo, tienen unos *loadings* no significativos en el gasto per cápita en salud, esto es, esta variable tiene menor importancia en éstos dos países y, no son los únicos.

En la *figura 7* los vectores correspondientes al imperio de la ley, la efectividad de gobierno, el control de corrupción y la calidad regulatoria tienen aproximadamente la misma longitud y sus ángulos no superan los 90°, indicando que estas variables están muy correlacionadas y siendo sus cargas muy similares. De esta forma, los países que aparecen alrededor de estas variables como Namibia, Sudáfrica, Senegal, Tunisia, Marruecos, Ghana, Bostwana, Cabo Verde etc. forman un grupo en donde destacan el imperio de la ley y el control de corrupción.

Figura 6. Biplot 2

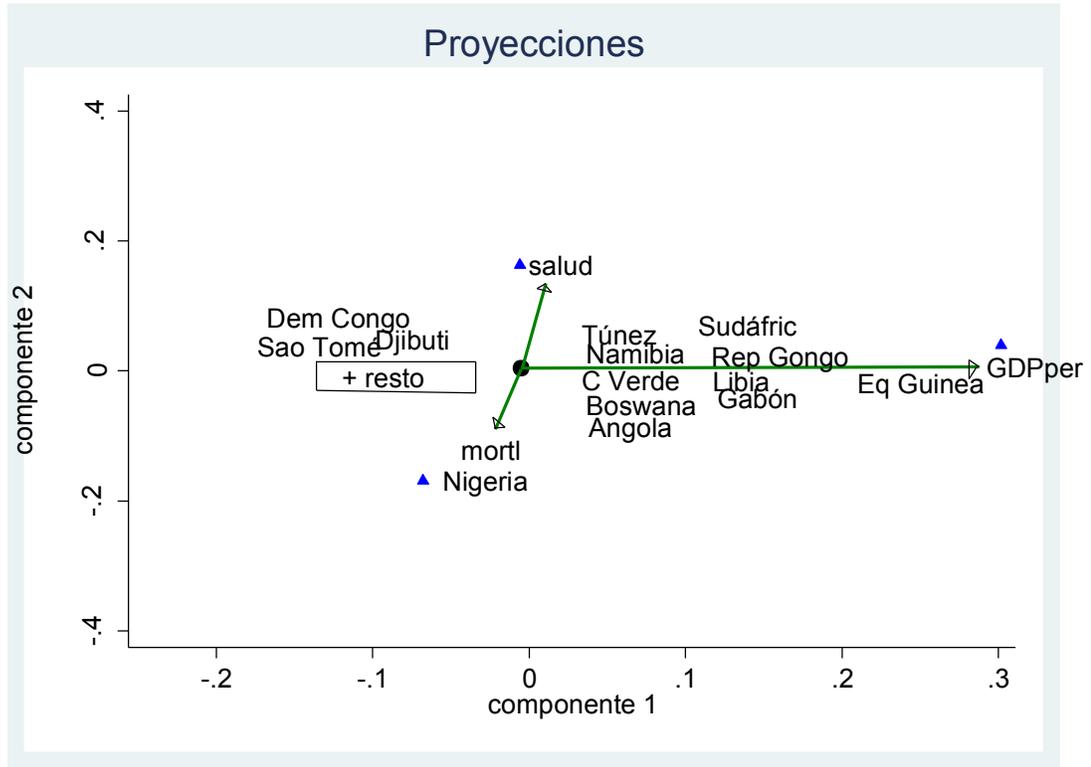
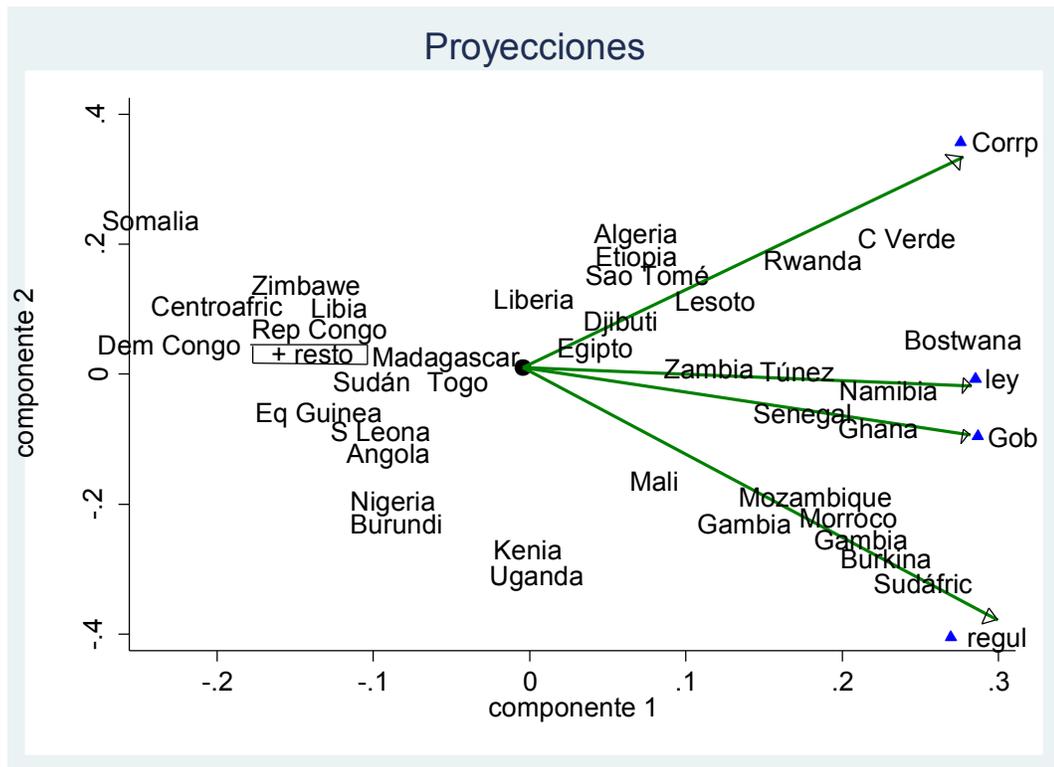


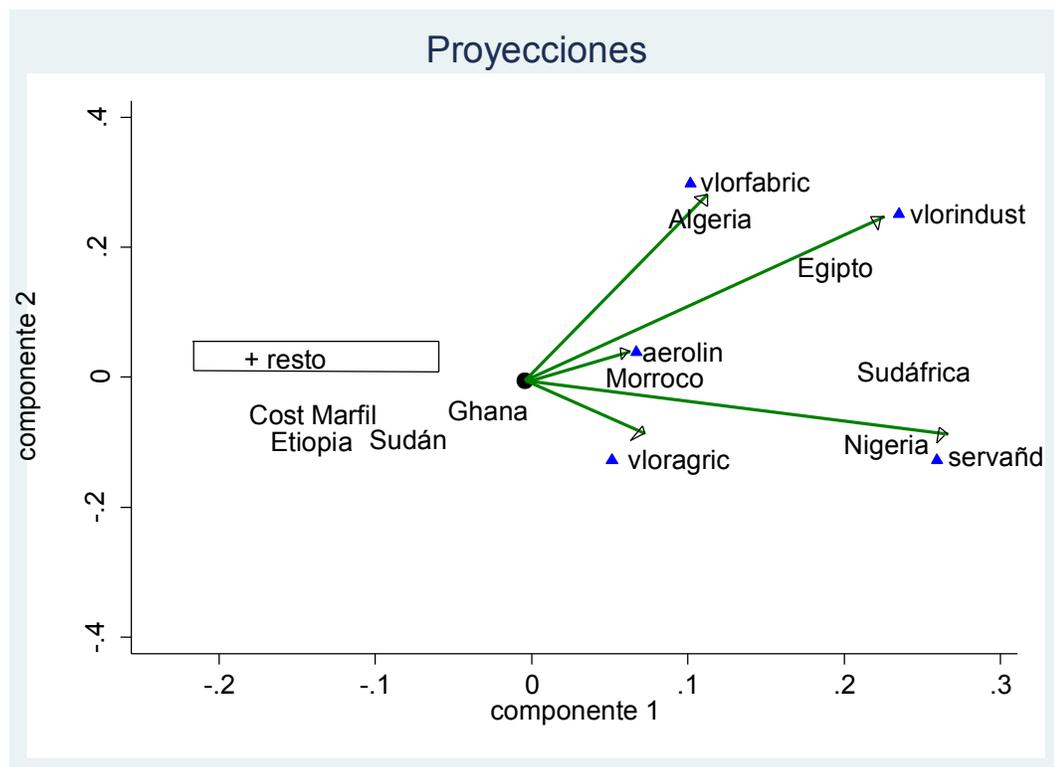
Figura 7. Biplot 3



Otra apreciación en la *figura 7* es la observada en su parte izquierda, son países situados a un ángulo de  $180^\circ$  de las flechas como son Liberia, Madagascar, Republica de Congo, Sao Tomé y Príncipe, Guinea Ecuatorial, Guinea, Centroáfrica etc. En estos países (que son la mayoría) predomina el complementario de la seguridad jurídica, esto es, la arbitrariedad, que es la actuación sin fundamento jurídico, no se garantiza el cumplimiento de la ley.

En la siguiente figura (*biplot 4*), se ilustra el grupo de países formado por Nigeria, Sudáfrica, Marruecos, Egipto y Argelia. Los ejes que corresponden a las variables valor añadido de la fabricación, valor añadido de la industria y servicios de valor añadido forman ángulos por debajo de los  $90^\circ$ , lo cual indica que estas variables están correlacionadas positivamente. Estas variables son destacables en el grupo de países anterior. Estos países reflejan el nivel tecnológico, el industrial y servicios alcanzado por África hasta hoy en día. El resto de países situado a  $180^\circ$  de los ejes, la mayoría, forma el grupo de países en los que no se destaca las características asociadas al desarrollo tecnológico e innovación y, éstos aparecen en la lista de los 49 países menos adelantados designados por las Naciones Unidas.

**Figura 8.** Biplot 4



Resumiendo todos los gráficos, se aprecia un patrón común en todos los *biplots* y es que la mayor concentración de los países se encuentra en el centro de los ejes, lo que nos hace entender que los países africanos a pesar de poseer características importantes que les hacen diferentes de los unos a los otros, pero sin embargo las magnitudes sociales, políticas e institucionales hacen de ellos países muy similares. Los factores sociales, políticos, históricos e institucionales son elementos comunes presentes en la mayoría de los países del continente y éstos, en gran medida, son los motivos por los que no es posible generar un crecimiento económico consolidado, capaz de resarcir el crecimiento demográfico. Por tanto, esta concentración es debido a que la mayoría de los países del continente tienen varias características comunes.

## 3.2 Análisis factorial

Vamos a comprobar si se verifican las condiciones para la aplicación del Análisis Factorial. Una posibilidad sería comprobar los supuestos de normalidad, linealidad y homoscedasticidad, ya que violaciones de estas hipótesis pueden disminuir las correlaciones entre las variables. Pero dado que nuestras variables se han estandarizado, esta posibilidad no es necesaria. Vamos a examinar la dimensión original realizando la factorización de la regla de Kaiser (*KMO*).

El estadístico *KMO* presenta un valor de 0.6766, superando el mínimo aceptable 0.5 y el test de esfericidad de Bartlett cuyo estadístico tiene un valor 1516.00 (*pvalue*  $\approx 0$ ), este test indica la presencia de correlaciones significativas. Todo ello nos indica que el espacio original de dimensiones puede ser apropiado para el Análisis Factorial.

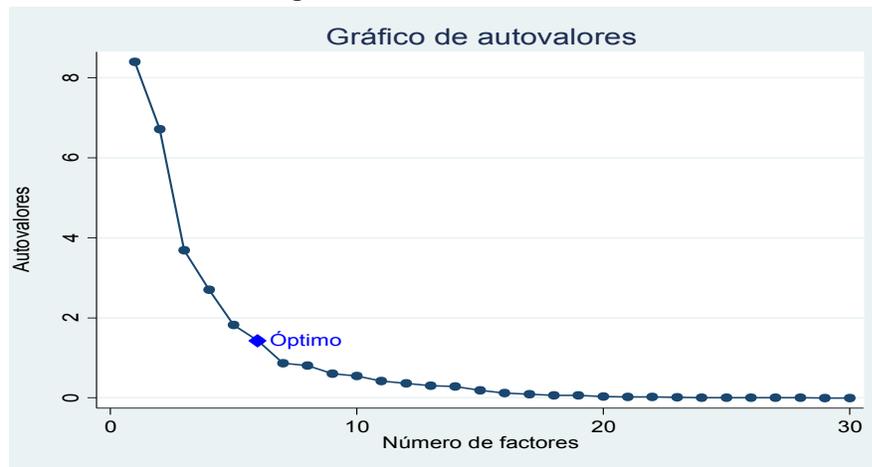
### 3.2.1 Extracción de factores

Para la extracción de los factores aplicamos el método de componentes principales. De acuerdo al criterio de Kaiser 6 factores serían suficientes y en total explicarían el 83.81% de la variabilidad total. Este resultado es confirmado por el gráfico de sedimentación (). Los autovalores expresan la cantidad de variabilidad explicada por cada factor. En la *tabla 4*, el primer factor explica un 28.44% de la varianza total, el segundo factor el 22.73%, el tercero lo hace en un 12.49% y el cuarto un 9.16%. Por tanto, los cuatro primeros factores explican casi el 73% de la variabilidad total de los datos. Podemos observar en la *figura 9* cómo los trozos a partir del factor siete empiezan a comportarse como unas líneas rectas, por lo que este criterio

indica retener hasta seis factores para explicar la estructura de las correlaciones entre las variables. Pero a pesar de las indicaciones de los dos criterios, en este trabajo retenemos los cuatro primeros factores para explicar las correlaciones entre las treinta variables originales. Las comunalidades nos indican qué cantidad de variabilidad de la variable implicada es explicada por los factores. Por ejemplo, el valor de la comunalidad de  $(1 - 0.1966) = 0.8034$  para la variable *desarrollo humano*, indica que tiene mucho en común con las otras variables, los cuatro factores explican el 80.34% de su variabilidad (*tabla 5*).

**Tabla 4.** Autovalores y variabilidad acumulada

Factor	Autovalor	Diferencia	Proporción	Acumulada
Factor1	8.39972	1.68681	0.2844	0.2844
Factor2	6.71291	3.02423	0.2273	0.5117
Factor3	3.68868	0.98436	0.1249	0.6366
Factor4	2.70432	0.87856	0.0916	0.7282
Factor5	1.82576	0.40498	0.0618	0.7900
Factor6	1.42077	0.56142	0.0481	0.8381
Factor7	0.85936	0.05498	0.0291	0.8672
Factor8	0.80438	0.20300	0.0272	0.8944
Factor9	0.60138	0.05496	0.0204	0.9148
Factor10	0.54642	0.12957	0.0185	0.9333
Factor11	0.41685	0.05757	0.0141	0.9474
Factor12	0.35928	0.05837	0.0122	0.9595
Factor13	0.30091	0.01428	0.0102	0.9697
Factor14	0.28663	0.09632	0.0097	0.9794
Factor15	0.19031	0.07525	0.0064	0.9859
Factor16	0.11506	0.02152	0.0039	0.9898
Factor17	0.09355	0.03124	0.0032	0.9929
Factor18	0.06231	0.00120	0.0021	0.9951
Factor19	0.06110	0.03030	0.0021	0.9971
Factor20	0.03080	0.00406	0.0010	0.9982
Factor21	0.02674	0.00491	0.0009	0.9991
Factor22	0.02183	0.01321	0.0007	0.9998
Factor23	0.00862	0.00288	0.0003	1.0001
Factor24	0.00574	0.00511	0.0002	1.0003
Factor25	0.00063	0.00036	0.0000	1.0003
Factor26	0.00027	0.00046	0.0000	1.0003
Factor27	-0.00020	0.00029	-0.0000	1.0003
Factor28	-0.00049	0.00217	-0.0000	1.0003
Factor29	-0.00266	0.00377	-0.0001	1.0002
Factor30	-0.00642	.	-0.0002	1.0000

**Figura 9.** Gráfico de sedimentación para los factores**Tabla 5.** Autovalores y especificidades

Variable	Factor1	Factor2	Factor3	Factor4	Unicidad
GDPper	0.7390	0.0479	0.4387	-0.1787	0.2272
HumDll	0.8218	0.0525	0.0548	-0.3495	0.1966
Expirt	0.0163	-0.1401	0.5877	-0.2690	0.5624
Desem	0.3946	0.0407	0.5592	-0.1981	0.4907
Ingr	-0.2768	0.2277	0.5510	-0.3998	0.4081
Bosque	-0.1525	-0.2639	0.4503	-0.0005	0.7043
Corrp	0.6814	-0.5076	-0.2153	0.0035	0.2317
polit	0.4487	-0.6718	0.2112	0.0867	0.2953
salud	0.8337	0.1264	0.3853	-0.1269	0.1243
VIH	0.3428	-0.1811	0.4923	0.4010	0.4465
tbrlsis	0.3227	0.0507	0.7304	0.3988	0.2007
mortl	-0.3677	0.2597	0.4234	0.6464	0.2002
costbusns	-0.6946	0.0446	0.0549	0.1249	0.4970
libInversn	0.3650	-0.4893	-0.2421	0.2463	0.5081
Agua	0.3743	-0.2686	-0.1085	0.1566	0.7514
ReserPetrol	0.1369	0.7967	-0.1438	0.1665	0.2982
OilProd	0.2601	0.8615	-0.1322	-0.0314	0.1716
Gob	0.8237	-0.4181	-0.1616	0.1801	0.0882
reserv	0.4031	0.5545	-0.0762	-0.2670	0.4530
EspVda	0.5212	-0.1465	-0.3969	-0.7269	0.0210
Deuda	-0.0379	-0.4319	0.0986	-0.1512	0.7794
ley	0.7685	-0.5238	-0.2101	0.1010	0.0807
regul	0.6575	-0.4461	-0.1389	0.3547	0.2237
aerolin	0.6073	0.3738	0.0127	0.0763	0.4855
vlorindust	0.5393	0.8175	-0.0461	0.0910	0.0305
vlorfabric	0.5253	0.7593	-0.1031	-0.1304	0.1199
servañd	0.5085	0.7514	0.0108	0.3369	0.0631
desmpjuv	0.5562	0.1019	0.4860	-0.1864	0.4093
vloragric	0.1484	0.8014	-0.2463	0.2550	0.2100
indinnov	0.6884	-0.2282	-0.3138	0.3485	0.2540

En cualquier caso, casi todas las variables tienen comunalidades por encima de 0.5 y, por tanto, es un buen indicio de solución factorial y procedemos a la rotación de los factores. La *tabla 5* muestra la matriz de cargas sin rotar. Las cargas o loadings indican las correlaciones entre las variables y cada uno de los factores.

### 3.2.2 Rotación de los factores

En la rotación, la comunalidad de cada variable no se modifica, esto es, la rotación no afecta a la bondad de ajuste de la solución factorial: aunque cambie la matriz factorial, las especificidades no cambian y por tanto, las comunalidades permanecen invariables. Sin embargo, cambia la variabilidad explicada por cada factor, luego los nuevos factores rotados no están ordenados de manera progresiva de acuerdo con la información que contienen, cuantificada a través de su autovalor o varianza. Las cargas rotadas por el método varimax aparecen en la *tabla 6*, mostrando sólo aquellas saturaciones cuyo valor absoluto sea superior a 0.4. Con objeto de conocer si la rotación ortogonal es la apropiada probamos también una rotación oblicua, mostrando sólo aquellos loadings cuyo valor absoluto sea superior a 0.4, *tabla 7*.

Las correlaciones entre los factores no superan en valor absoluto 0.3. En consecuencia, los resultados obtenidos con la rotación varimax son los que mantendremos, aunque en nuestro caso, los resultados no tienen diferencias significativas.

Tabla 6. Rotación varimax

Variable	Factor1	Factor2	Factor3	Factor4	Unidades
GDPper			0.8045		0.2272
HumD11			0.6038	0.4502	0.1966
Exprt			0.5409		0.5624
Desem			0.7133		0.4907
Ingr		-0.6459	0.4036		0.4081
Bosque					0.7043
Corrp		0.7963			0.2317
polit	-0.4557	0.6047			0.2953
salud			0.8028		0.1243
VIH			0.4443	-0.4898	0.4465
tbrlsis			0.6275	-0.6204	0.2007
mortl				-0.8473	0.2002
costbusns		-0.4811			0.4970
libInversn		0.6704			0.5081
Agua		0.4922			0.7514
ReserPetrol	0.8145				0.2982
OilProd	0.8792				0.1716
Gob		0.9089			0.0882
reserv	0.6077				0.4530
EspVda				0.9307	0.0210
Deuda	-0.4439				0.7794
ley		0.9037			0.0807
regul		0.8678			0.2237
aerolin	0.5665				0.4855
vlorindust	0.9468				0.0305
vlorfabric	0.8642				0.1199
servañd	0.9080				0.0631
desmpjuv			0.7446		0.4093
vloragric	0.8636				0.2100
indinnov		0.8404			0.2540

Tabla 7. Rotación promax

Variable	Factor1	Factor2	Factor3	Factor4	Unicidad
GDPper			0.8018		0.2272
HumDll			0.5368		0.1966
Exprt			0.6570		0.5624
Desem			0.7742		0.4907
Ingr	-0.7856		0.5585		0.4081
Bosque		-0.4212			0.7043
Corrp	0.7555				0.2317
polit	0.5508	-0.4898			0.2953
salud			0.7761		0.1243
VIH			0.4596	0.5479	0.4465
tbrlsis			0.6864	0.6827	0.2007
mortl				0.8607	0.2002
costbusns					0.4970
libInversn	0.7238				0.5081
Agua	0.5045				0.7514
ReserPetrol		0.8367			0.2982
OilProd		0.8758			0.1716
Gob	0.8864				0.0882
reserv		0.5691			0.4530
EspVda				-0.9344	0.0210
Deuda		-0.4650			0.7794
ley	0.8757				0.0807
regul	0.9009				0.2237
aerolin		0.5419			0.4855
vlorindust		0.9336			0.0305
vlorfabric		0.8347			0.1199
servañd		0.9151			0.0631
desmpjuv			0.7727		0.4093
vloragric		0.9006			0.2100
indinnov	0.8916				0.2540

El objetivo principal del análisis factorial consiste en identificar la estructura de las relaciones entre las variables, la cual viene definida por los factores, factores que podemos identificar de la siguiente manera:

*Factor 1:* viene ligado con las variables *reservas de petróleo, producción de petróleo, reservas de divisas, valor añadido de la fabricación, valor añadido de la industria, valor añadido de la agricultura, servicios de valor añadido, pasajeros de aerolíneas y deuda extranjera*. Este factor puede ser etiquetado como *petrodólares*. Las cargas rotadas relacionadas con estas variables tienen signo positivo, además de la deuda externa que tiene una carga negativa. Esto, indica las disputas por el control de las reservas de petróleo, lo que disminuye la estabilidad política y

debido al aumento de los ingresos procedentes de la explotación del petróleo disminuye la deuda extranjera. El aumento de ingresos que dinamiza la economía hacen mejorar las demás variables como es valor añadido de la fabricación, el valor añadido de la industria, el valor añadido de la agricultura, los servicios de valor añadido y los pasajeros de aerolíneas. La explotación del petróleo se puede incluir en los tres sectores de la economía según la fase que se considere. Es un recurso natural que se explota (sector primario), la perforación y la refinería (secundario), la venta, las exportaciones, y el transporte (servicios). En síntesis, la explotación del petróleo aporta valor añadido a muchos sectores y los ingresos obtenidos ayudan a invertir en otros sectores como la agricultura.

*Factor 2:* está asociado con las variables *control de corrupción, estabilidad política, libertad de inversión, coste de iniciar un negocio, acceso al agua potable, efectividad de gobierno, imperio de la ley, calidad regulatoria, el índice de innovación e ingresos de los recursos naturales*. Este factor puede ser identificado con **la seguridad jurídica y garantía de hacer negocios**. Las cargas rotadas asociadas a las variables contenidas en este factor tienen el mismo signo, menos las variables ingresos de los recursos naturales y el coste de iniciar un negocio que tienen cargas con valores negativos. Esto indica que, tanto la efectividad gubernamental, la estabilidad política, el control de la corrupción e imperio de la ley pueden, en cierto grado, instaurar como una obligación a los estados, garantizar a todas las personas el acceso al agua potable para el uso personal y doméstico. Este factor está asociado con los países Bostwana, Namibia, Ruanda, Cabo Verde, Gabón y otros. Todo lo anterior repercute en la mejora de la gestión económica, reduce el coste de iniciar un negocio. Al reducir este coste, habría una mayor inversión, muchos negocios. Una mayor inversión significa un aumento y mejora de la innovación y, por tanto, la economía se industrializa. La economía aumenta el peso de los sectores secundario y terciario y, se reduce la importancia del sector primario, lo cual, por consiguiente, reduce los ingresos de los recursos naturales.

*Factor 3:* viene asociado con las variables *renta per cápita, desarrollo humano, exportaciones de bienes, ingresos de recursos naturales, tasa de desempleo, desempleo juvenil, gasto per cápita en salud, infectados de VIH e infectados de tuberculosis*. Este factor puede ser descrito como representativo del **crecimiento económico y poca formación**. Todas las cargas rotadas relacionadas con estas variables tienen signo positivo. Es decir, estos países africanos experimentan un crecimiento económico que se traduce en un aumento del PIB per cápita, así

como en un aumento en la inversión en salud pero por los altos índices de analfabetismo o bajo nivel de educación, aumenta el nivel de desempleo en la población. El crecimiento económico provoca grandes desplazamientos de la población y grandes aglomeraciones en las grandes ciudades, lo que provoca que todos no sean absorbidos en el mercado laboral. Estas aglomeraciones y poca formación también provocan el aumento de las infecciones de las enfermedades, tanto de transmisión sexual y/o no. Al tener poca formación, la mayor parte de esta población no suele encontrar empleo y toman muy pocas precauciones para evitar los contagios de estas enfermedades. Por otro lado, en muchos países africanos, las culturas y los prejuicios sociales suelen jugar un papel desfavorable en cuanto a tomar medidas de protección sexual. En muchos casos, hacen falta programas de educación sexual y sensibilización a la población para evitar los aumentos de contagios. En definitiva, aunque muchos países mejoran económicamente, el nivel de educación y formación de la población (y los motivos culturales en muchos casos) hace que muchos estén desempleados y que aumenten los contagios de enfermedades de contagio como el sida, la sífilis y la tuberculosis.

*Factor 4:* está relacionado con las variables *desarrollo humano, esperanza de vida, índice de mortalidad, infectados de VIH e infectados de tuberculosis*. Este factor se puede identificar con **el desarrollo económico**. Las cargas rotadas asociadas a las variables desarrollo humano y esperanza de vida tienen signo positivo y negativo el índice de mortalidad, los infectados de VIH e infectados de tuberculosis. Lo cual indica que, el desarrollo económico mejora el nivel de vida de las personas, lo que influye positivamente en su esperanza de vida. Reduce la mortalidad y también repercute en el desarrollo humano que mejora, precisamente, con un aumento de los ingresos, mejora del acceso a la sanidad y el nivel de educación.

### 3.2.3 Validación factorial

El objetivo es garantizar la consistencia interna de los factores, es decir, vamos a corroborar la fiabilidad de los factores. Vemos si los factores miden realmente las correlaciones de las variables. Para ello hacemos el test de Crombach, lo ideal es que esté por encima de 0.7 pero también es aceptable a partir del 0.5. Este coeficiente estima la media de todos los coeficientes de correlación de las variables dentro del factor.

El alpha de Crombach para el factor 1 proporciona un valor de 0.56, el del factor 2 de 0.74, el del factor 3 de 0.63 y el del factor 4 de 0.8833. Únicamente el coeficiente asociado al primer

factor está muy cerca del valor mínimo aceptable. Sería deseable que futuras investigaciones sugiriesen nuevas variables para este factor y le proporcionasen todavía una mayor fiabilidad. En cualquier caso, todas las variables son fiables y todos los factores evidencian una alta consistencia interna.

### 3.3 Análisis gráfico de los conglomerados

En esta sección del trabajo, el objetivo es proyectar grupos de países africanos con características similares mediante figuras en forma de mapas y dendrograma.

**Figura 10.** Mapa 2

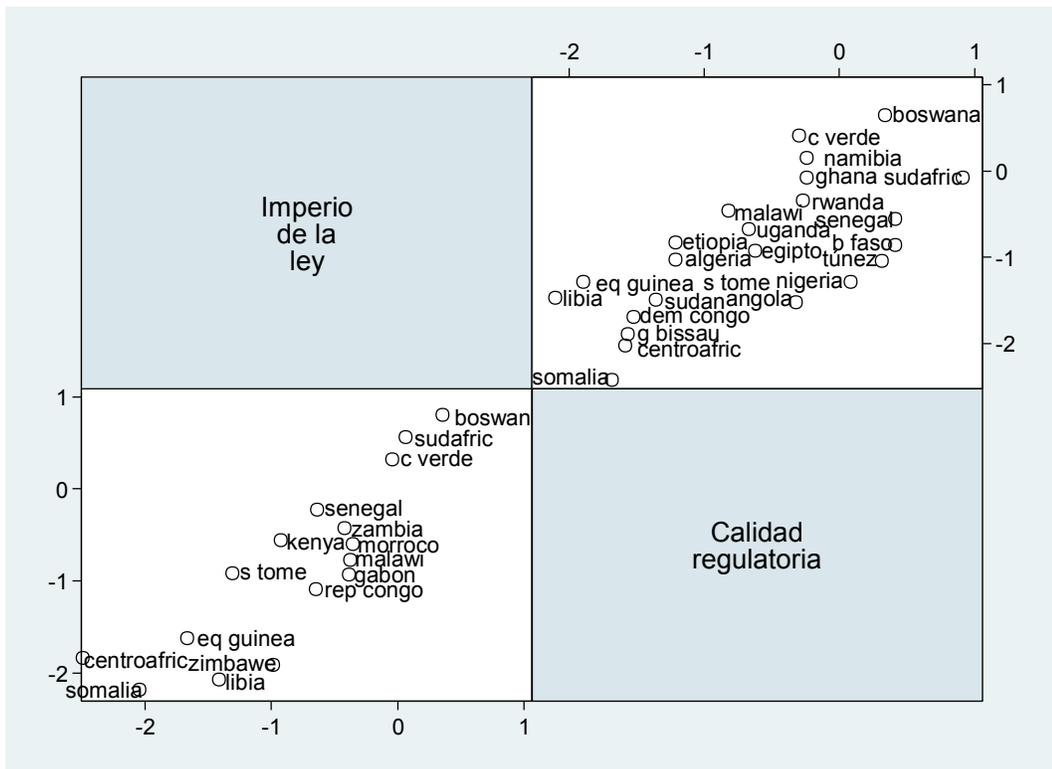




Figura 13. Mapa 5

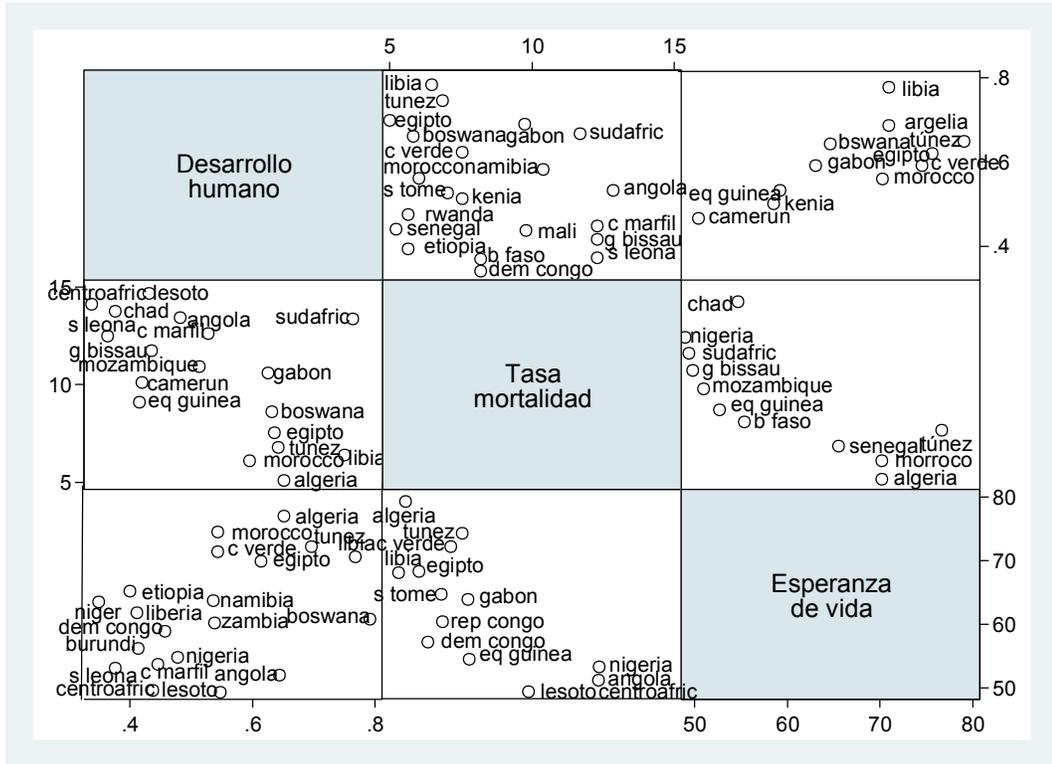
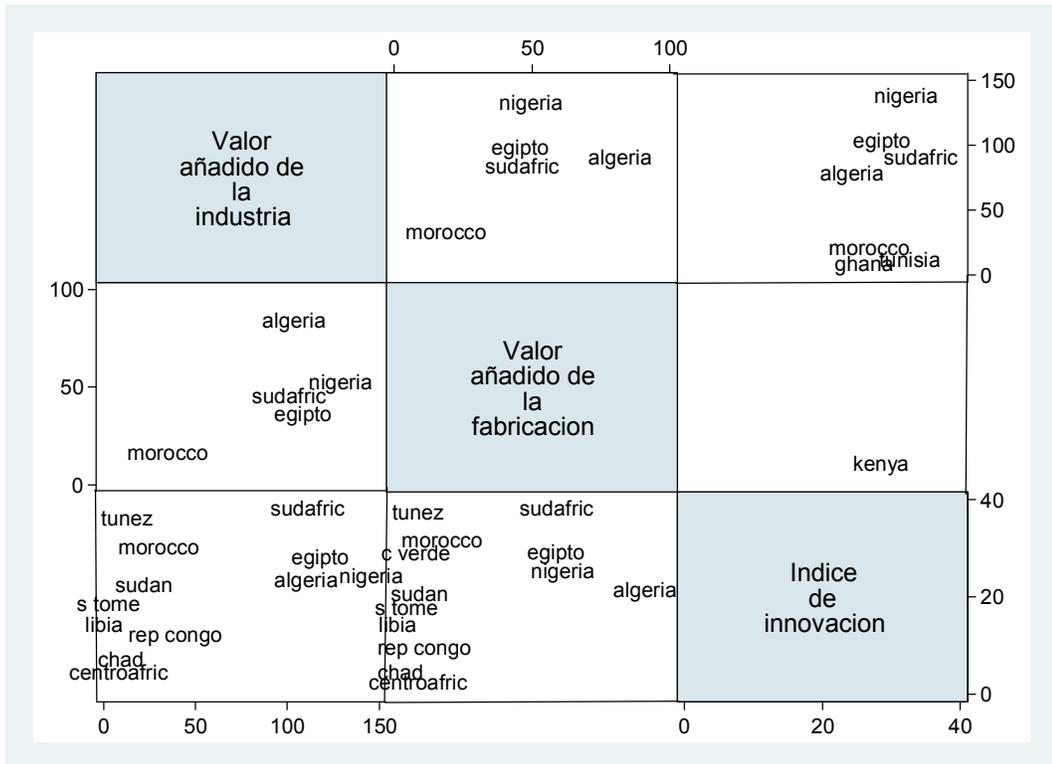


Figura 14. Mapa 6

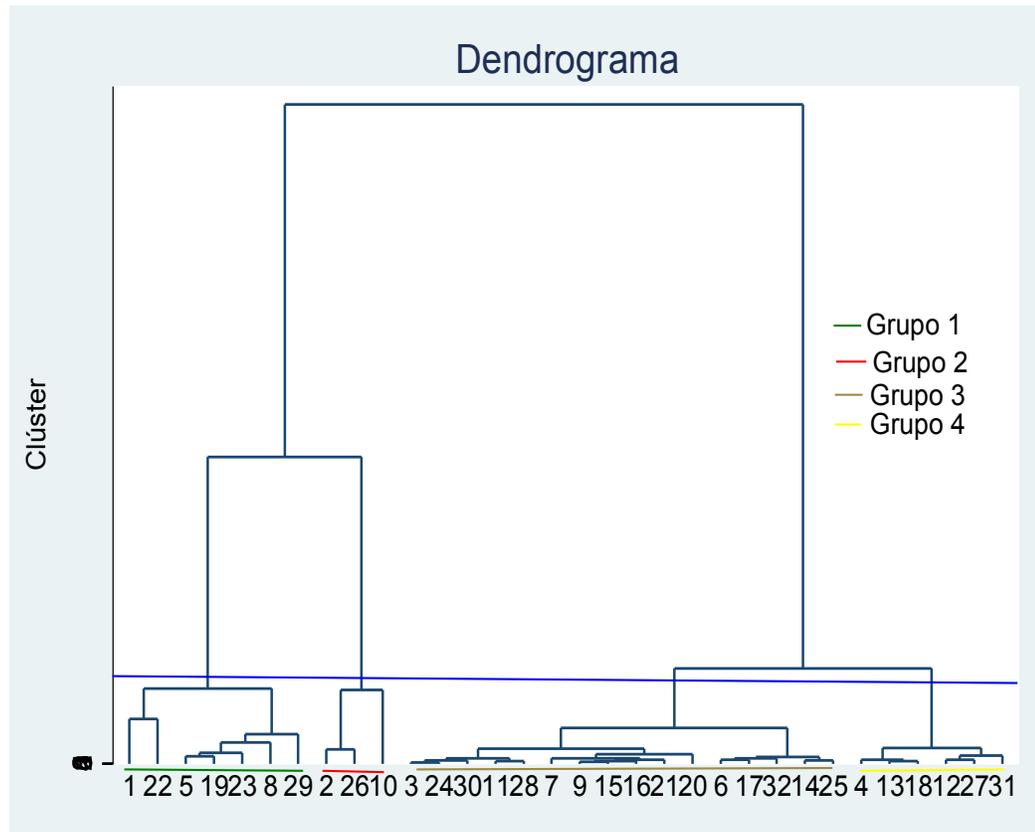


Tres características importantes tienen los mapas anteriores: las posiciones atípicas entre las observaciones en el cuadrante representan el nivel de dominio de las  $n$  variables sobre el grupo de ítems correspondiente, las posiciones indican también el grado de similitud o heterogeneidad entre los países y los patrones de los ítems en cada cuadrante representan aproximadamente la correlación entre las variables.

Diversas características son detectadas: en el *mapa 2*, las observaciones correspondientes a Bostwana, Sudáfrica, Tunisia, Namibia y Cabo Verde están localizadas en la parte superior derecha. Son aquellos países en los que predomina el marco jurídico; mientras que en la parte inferior izquierda sitúa a los países como Somalia, Centroáfrica, Libia, Guinea Ecuatorial, Rep. De Congo, Zimbawe y otros, que formarían el grupo de países en el que no se garantiza la ley. El *mapa 3* sitúa a Nigeria, Libia, Angola, Algeria y Egipto en un grupo en el que predomina el sector petrolero y son los países que pierden la estabilidad política debido a las disputas por el control de las reservas de petróleo; por esto, todos estos países se sitúan en la parte inferior de la zona correspondiente a la estabilidad política. El *mapa 4* agrupa los países Guinea Ecuatorial, Gabón, Libia, Rep. De Congo y Algeria como grupo en el cual el sector de recursos naturales es capital; dicho grupo lo conforman los países del continente que experimentan el *síndrome holandés*, esto es, a pesar del crecimiento económico de estos países debido a la explotación y exportación de sus recursos naturales, sus índices de desempleo son elevados por la pérdida de la competitividad en la industria local. En el *mapa 5*, las variables *desarrollo humano* y *esperanza de vida* presentan un patrón lineal de carácter directo y su cuadrante de fusión sitúa en la parte superior derecho a Libia, Tunisia, Marruecos, Cabo Verde, Namibia, Bostwana y Sudáfrica y, se sitúan en la parte inferior izquierda en los cuadrantes correspondientes a la *tasa de mortalidad*, lo que indica que, este grupo dominado por África del norte o septentrional se caracteriza por la importancia del *desarrollo de humano y la esperanza de vida*. En el *mapa 6* se aprecia que casi todos los países que aparecen se caracterizan por tener una cultura emprendedora, de hecho, en el mapa, casi aparece la minoría del conjunto de países africanos. Entendiendo que, en el continente no destaca una cultura emprendedora en la mayoría de los países.

Como se ha dicho anteriormente, el análisis clúster trata de identificar los ítems vectoriales que son similares y agruparlas entonces en clústers o grupos. Una proyección de esta clasificación es el siguiente dendrograma, obtenido mediante el procedimiento jerárquico Ward. En él, cortando la línea azul, se ha quedado con cuatro grupos de países.

**Figura 15.** Dendrograma

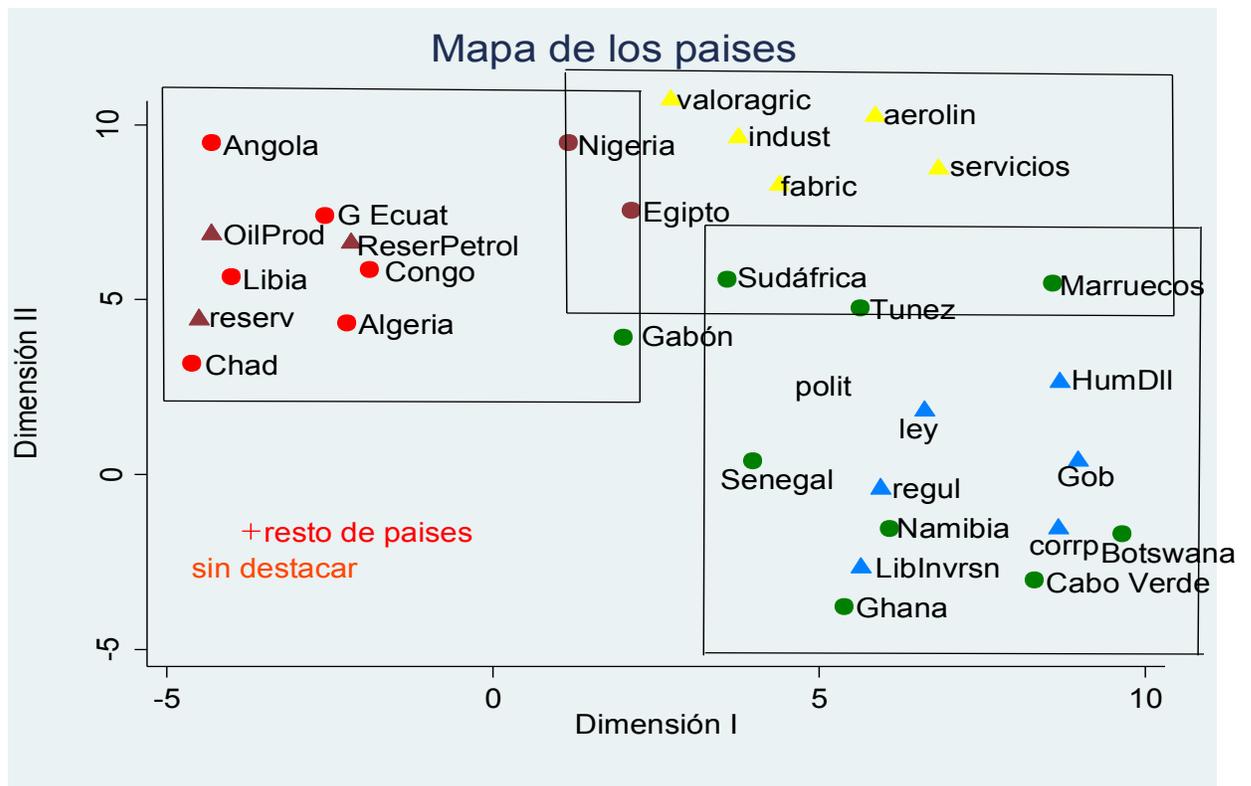


El dendrograma establece una estructura de grupos en la que los países dentro de cada clúster son lo más parecido posible, pero la diferencia entre grupos es la máxima posible. El primer grupo está formado por los países Lesoto, Burundi, Guinea Bissau, Liberia, Centroáfrica, etc. Este grupo está muy caracterizado por la pobreza, peor calidad de regulación y la arbitrariedad. El segundo grupo, formado por Angola, Malawi, Rep. De Congo, etc. Este grupo destaca por los elevados ingresos de recursos naturales, porcentaje de tierra forestal, sin estabilidad política ni control de corrupción, elevado índice de mortalidad y el alto coste a la hora de iniciar negocio. Seguramente son de los países con muchas disputas internas por el control de las reservas naturales, lo cual disminuye la estabilidad política. El tercer grupo está formado por Bostwana, Libia, Cabo verde, Gabón, Gambia, Costa de Marfil, Kenia, Djibouti, Chad, Mozambique, Ghana, Etiopia, Camerún, etc. En este grupo predomina la elevada deuda externa, el desempleo, el peor nivel de industrialización, la ganadería etc. Pero sorprendentemente presentan además, un alto ratio de esperanza de vida y un destacable desarrollo humano. El cuarto grupo, formado por Burkina-Faso, Egipto, Guinea Ecuatorial, Mali, Namibia, Guinea Conakry, etc. Este grupo

destaca por sus elevados ingresos de divisas, ingresos de recursos naturales, el gasto de salud per cápita, dificultad a la hora de crear un negocio y sin calidad de regulación.

Los países africanos a pesar de poseer características importantes que les hacen diferentes de los unos a los otros, sin embargo, las magnitudes sociales, políticas, históricas e institucionales hacen de ellos países bastante similares. La figura siguiente (*mapa 7*), hace una proyección de grupos de países clasificados en función de las variables que más diferente hacen a los países de los otros dentro del continente.

**Figura 16.** Mapa 7



Como podemos apreciar en la figura anterior, la clasificación está bastante evidente. Los países situados en la parte inferior derecha destacan por una seguridad jurídica y garantía de hacer negocios. Los situados en la parte superior derecha se caracterizan por poseer un importante valor añadido de la industria y fabricación. Los países ubicados en la parte superior izquierda son aquellos en los que predomina el sector petrolero, mientras que los situados en la parte inferior izquierda forman el conjunto mayoría en el que no destacan las características anteriores, éstos últimos serían de los más pobres del continente.

## 4 Conclusiones y potenciales extensiones

Aun reconociendo la importancia de la positiva evolución mostrada por la exportación de recursos naturales y su papel de impulsor de la positiva senda iniciada en la pasada década por un buen número de países africanos, no debemos obviar el papel que han jugado elementos tales como el fin de algunos conflictos armados –y la posterior estabilidad política necesaria para impulsar el crecimiento–, que la mejora de los fundamentos macroeconómicos y el control de sus variables clave, y que las reformas emprendidas para generar un mejor clima para el desarrollo de los negocios. No sería atrevido decir que todos estos factores y en especial los cambios estructurales, han contribuido a generar un auténtico cambio en términos de productividad, en el que las empresas han crecido en escala, han aumentado su inversión y se han convertido en más competitivas.

Sin embargo, el que esta senda de crecimiento iniciada se convierta en estable y no ocurra como en experiencias previas fallidas, este crecimiento ha de tener unas bases sólidas y no ser un proceso volátil sostenido tan sólo por una evolución favorable de precios y demanda de recursos naturales. Para evitar episodios ya conocidos, estas economías han de aprovechar períodos expansivos como los iniciados en este siglo, para acometer cambios estructurales profundos y disminuir su exposición a las evoluciones de la demanda internacional y encontrar nuevos posicionamientos que les permitan aumentar su poder de negociación y evitar la dependencia de ciertos sectores, empresas y países. Sin embargo, hay que ser conscientes de que la realidad africana es muy diversa y que aunque las perspectivas a nivel global han de ser consideradas positivas, las trayectorias de crecimiento de cada país son y serán muy diversas.

Atendiendo a su nivel de diversificación productiva, en el grupo de países del continente cabe distinguir un grupo de economías diversificadas, entre las que se encuentran Egipto, Marruecos, Sudáfrica y Túnez (las más ricas y con PIB menos volátil ya que el consumo interno es el principal motor de su demanda y sus déficit externos son los más bajos. Avanzan hacia la producción de productos de mayor valor añadido y sus sistemas educativos y sanitarios avanzan de manera decidida. Otro grupo se encuentra compuesto por los exportadores de petróleo y gas, que aunque tienen los PIB per cápita más altos tienen las economías menos diversificadas (Argelia, Angola, Chad, Congo, Guinea Ecuatorial, Gabón, Libia y Nigeria) y cuya crecimiento económico depende, en última instancia– de la evolución de la demanda y de los precios. Estos

ingresos les permiten cubrir sus déficit aunque algunos de estos países han iniciado su transición a transformarse en economías más diversificadas. Por su parte las llamadas economías en transición, son aquellas que tienen PIB per cápita más bajos que los dos anteriores grupos, un grupo de países compuestos por Camerún, Ghana, Kenya, Mozambique, Senegal, Tanzania, Uganda, y Zambia han iniciado el proceso de diversificación de sus economías que son muy dependientes de la producción de algún recurso o materia prima y en las que la agricultura es aún un sector que representa un alto porcentaje del PIB, aunque tratan de aumentar sus exportaciones de manufacturas.

Nuestro análisis, ha permitido detectar en base aun mayor número de dimensiones, una estructura latente de factores que permite discriminar/agrupar a estos países y que hemos convenido en interpretar en la disponibilidad de petróleo o gas (petrodólares), el grupo de países en los que existe una seguridad jurídica y un entorno favorable para el desarrollo de los negocios, los más avanzados en términos de crecimiento y los que disponen de peores dotaciones de capital humano, y que son los que marcarán no sólo el posible establecimiento de estrategias conjuntas sino el participar de unos mismos clubs de convergencia, esto es, de condiciones para alcanzar una cierto *catch up* entre ellos y experimentar experiencias de crecimiento parecidas.

Explorar cómo disponer de los fundamentos, dotaciones de factores y características de los que están evolucionando de forma más adecuada y la expansión del comercio entre los propios países africanos ha de ser una de las claves que contribuya al desarrollo de todos estos países, especialmente al de los dos últimos grupos.

Esta estrategia debe ser común a todos ellos, que son demasiado pequeños para acometer estrategias individuales de desarrollo pero a los que la integración regional, las mejoras en las infraestructuras y los nuevos sistemas regulatorios podrían permitirles acceder a mercados mayores como paso previo para competir internacionalmente con otras economías emergentes de bajos costes de producción.

## Potenciales extensiones

Aunque en este trabajo nos hemos limitado a aplicar técnicas de análisis multivariante con fines exploratorios, somos conscientes que la extensión natural de este trabajo ha de pasar por la propuesta metodológica y de obtención de algún tipo de indicador sintético, en el que este

análisis exploratorio no sea más que una primera parte del proceso de selección de indicadores, y en el que el análisis inferencial complemente a este primer análisis exploratorio. Todo ello con el fin de poder monitorizar y predecir cambios de individuos entregrupos e intragrupos. De la misma manera, su uso para contrastar hipótesis acerca de la explicar las sendas de crecimiento experimentadas por estos grupos de países y las posible existencia de clubs de convergencia, se torna como la evolución más lógica de esta línea de trabajo.

Por otra parte, una segunda limitación del trabajo viene dada por el uso de técnicas de análisis paramétrico, esto es, técnicas basadas en supuestos de cuyo cumplimiento depende la validez de los resultados. Tanto por razones de búsqueda de robustez como para evitar estos problemas de potencial incumplimiento de los supuestos, el análisis anterior debería ser completado/refrendado con un análisis de robustez haciendo uso de técnicas de análisis de datos no supervisados, esto es con datos de entrenamiento. Así, métodos no paramétricos en los que no ha de verificarse ningún supuesto para su correcta utilización como el del vecino más cercano, los basados en máquinas de soporte vectorial o el de los arboles de decisión pueden ser utilizados para este fin.

Por otro lado, el intento de mejorar el número de indicadores y la calidad de los datos y series incorporados en el mismo ha de ser también tarea esencial a la hora de completar este ensayo.

Sin embargo, quizás sea la detección de la influencia con la que los diferentes fundamentos de estas economías ayudan a explicar la extrema diversidad en evoluciones y ritmos de crecimiento, sea el principal reto y la deriva más interesante de este primer trabajo sobre las economías del continente africano.

## 5. Referencias

- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests, *Proceedings of the Royal Society of London Series A*, 160, 268–282.
- Bartlett, M.S. (1954). A Note on the Multiplying Factors for Various Chi Square Approximations. *Journal of the Royal Statistical Society*, 16 (Series B), 296–298.
- Barro, R.J., 1991. Economic growth in a cross section of countries. *Quarterly Journal of Economics* 106,407–444.
- Bertocchi, G., & Canova, F. (2002). Did colonization matter for growth? An empirical exploration into the historical causes of Africa's underdevelopment. *European Economic Review* 46, 1851–1871.
- Cronbach, L.J. (1951). Coefficient Alpha and the Internal Structure of Tests, *Psychometrika*, 16, 297–334.
- Easterly, W., Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics* 112, 1203–1250.
- Gnanadesikan, R. (1997). *Methods for Statistical Data Analysis of Multivariate Observations*. New Jersey: Rutgers University New Brunswick.
- Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*, MIT Press.
- Harris, C.W. and Kaiser, H.F. (1964). Oblique Factor Analytic Solutions by Orthogonal Transformation, *Psychometrika*, 32, 363–379.
- Hartigan, J.A. (1981). Consistence of Single Linkage for High-Density Clusters, *Journal of the American Statistical Association*, 76, 388–394.
- Hastie, T., Tibshirani, R., and Friedman, J.H.(2009). *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, New York: Springer Science and Business Media.
- Jollie, I. (1986). *Principal Component Analysis*. Springer-Verlag, New York, 1986.
- Levine, R., Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82, 942–963.
- Rencher, A. (2002). *Methods of Multivariate Analysis*, Brigham Young University: Second Edition.

- Rodrik, D., 1998. Trade policy and economic performance in Sub-Saharan Africa. *NBER Working paper* No. 6562.
- Sachs, J., Warner, A. (1997). Sources of slow growth in African economies. *Journal of African Economies* 6, 335–376.
- Schmidt-Hebbel, K. (1996). Fiscal adjustment and growth: In and out of Africa. *Journal of African Economies* 5 (Suppl. Part I), 7–59.