

# **Estudio Exploratorio: Análisis de Componentes Principales y de Cluster en RStudio**

Florencia Mosconi

Licenciatura en Economía

Universidad Nacional de Villa Mercedes

Dr. Cristian Rabanal

3 de diciembre de 2021

## Tabla de contenido

<b><i>Introducción</i></b> .....	<b>3</b>
<b><i>Marco teórico</i></b> .....	<b>6</b>
<b><i>Datos y metodología</i></b> .....	<b>7</b>
<b>Tabla de Datos</b> .....	<b>7</b>
<b>Metodología</b> .....	<b>11</b>
Análisis de Componentes Principales .....	11
Análisis de Cluster .....	12
<b><i>Resultados</i></b> .....	<b>13</b>
<b>Tabla de Datos</b> .....	<b>13</b>
Análisis de la Varianza .....	14
<b>Análisis de Componentes Principales</b> .....	<b>14</b>
Construcción y análisis de la matriz de correlación y test KMO .....	14
Selección de Componentes .....	18
Matriz de Rotación .....	21
Interpretación de los Componentes y Caracterización de Individuos .....	22
Construcción de las Nuevas Variables .....	25
<b>Análisis de Cluster</b> .....	<b>26</b>
Análisis de Tendencia de Agrupación .....	26
Determinación del Número de Clusters .....	28
Método K-medias e Interpretación de Resultados .....	33
<b><i>Consideraciones Finales</i></b> .....	<b>41</b>
<b><i>Referencias</i></b> .....	<b>42</b>
<b><i>Anexos</i></b> .....	<b>43</b>

## **Introducción**

El objetivo del presente trabajo es proveer un estudio exploratorio de dos de los métodos multivariantes más utilizados para la interpretación de información, para ello se agrupará a los países analizados en el Programa de las Naciones Unidas para el Desarrollo (PNUD) según sus semejanzas correspondientes a las siguientes dimensiones, educación, economía, salud, ambiente, desigualdad y sociedad.

Para el armado de la matriz de datos se recurrirá al Programa de las Naciones Unidas para el Desarrollo (PNDU), que provee información de un amplio número de variables para 195 países, de los cuales se seleccionará sólo a aquellos que contengan datos para todas las variables escogidas.

Se aplicará el método factorial de Análisis de Componentes Principales para lograr reducir la dimensionalidad de los datos, facilitando el manejo de información y conservando la máxima variabilidad presente en los valores originales, destacando que, cuanto mayor es la varianza de los datos mayor es la información analizada. A través del Análisis de Cluster se realizará el armado de grupos de individuos que presenten semejanzas en las variables analizadas, para luego proceder a la caracterización de cada grupo y su representación gráfica.

Los objetivos que se buscarán concretar con la realización de este estudio serán:

1. Obtener grupos de países por grados de asociación de las variables analizadas correspondientes a las dimensiones de salud, educación, economía, desigualdad, sociedad y medio ambiente.

- 1.1. Aplicar el método de Análisis de Componentes Principales (ACP) reduciendo la dimensionalidad de la tabla de datos construida para, a través de nuevas

variables, realizar el análisis de los atributos de los individuos de manera simple y mediante representación gráfica

- 1.2. Construir la matriz de correlación para evaluar la fuerza y dirección de la relación entre las variables tipificadas seleccionadas y realizar las pruebas de aleatoriedad a través del coeficiente de Pearson
- 1.3. Aplicar el Test de Kaiser, Meyer y Olkin para determinar si nuestros datos son susceptibles al análisis de componentes principales
- 1.4. Realizar pruebas estadísticas mediante herramental gráfico para determinar el número de componentes óptimo que maximicen la varianza explicada
- 1.5. Calcular los componentes principales para obtener la conformación de cada nueva variable independiente
- 1.6. Interpretar las características de las observaciones que opone cada componente y definir su función
- 1.7. Explorar los atributos de al menos dos países en función de su ubicación en los componentes principales obtenidos
- 1.8. Realizar un Análisis de Cluster sobre nuestra tabla de datos para agrupar a los países seleccionados en función de las variables
- 1.9. Construir la matriz de distancia euclídea para observar tendencias entre las unidades de observación
- 1.10. Aplicar el Test de Hopkins para determinar si nuestros datos son agrupables

- 1.11. Analizar pruebas estadísticas de selección y su análisis gráfico para determinar la cantidad óptima de cluster a escoger
- 1.12. Definir la composición de cada cluster según el peso de cada variable en cada uno de ellos para caracterizar los países que los componen
- 1.13. Demostrar gráficamente la composición y ubicación bidimensional de componentes principales, los cluster contruidos

### **Marco teórico**

En la actualidad “en la actividad económica moderna, los datos se constituyen como un factor esencial para la producción, tal como los activos fijos y el capital humano” (Rodríguez, Palomino, & Mondaca, 2017). El análisis de componentes principales y de clúster son dos métodos multivariantes muy utilizados para el análisis de datos masivos y para la creación de indicadores compuestos.

El análisis de componentes principales (ACP) es un método algebraico/estadístico que busca simplificar, reducir y estructurar la información contenida en una matriz de datos (Colina & Roldán, 1991). La principal fortaleza de este método es resumir una serie de indicadores individuales preservando la máxima proporción posible de la variación de la información original (OECD, 2008).

Esta técnica es considerada de análisis multivariable, ya que estudia y permite interpretar el efecto simultáneo de múltiples variables de cada unidad de observación. A través de la transformación de variables correlacionadas en un nuevo conjunto de variables independientes, llamadas componentes principales (Jolliffe, 2002), este método reduce la dimensionalidad permitiendo interpretar de forma más sencilla y GGG una amplia cantidad de información.

Los objetivos del ACP se pueden resumir en dos, por un lado, ‘comparar los individuos con la información disponible para evaluar semejanza a través de los caracteres considerados buscando grupos de individuos que presenten valores semejantes para hallar tipologías de individuos’, y por otro, ‘evaluar relación existente entre las variables buscando grupos de variables correlacionadas para hallar tipologías de variables’ (Baronio & Vianco, 2014).

El segundo análisis que realizaremos será el de Cluster, que es utilizado para agrupar información sobre observaciones en función de su similitud en diversos indicadores individuales. Éste método nos permitirá mediante la creación de indicadores compuestos conglomerar los países de nuestra tabla de datos y simplificar su caracterización y comparación (OECD, 2008).

Este método exploratorio multivariante es aplicado en diferentes disciplinas, como la bioinformática, el marketing y la segmentación de imágenes, al permitir analizar información masiva provenientes de un amplio número de variables y unidades de observación.

Es un método de aprendizaje no supervisado, es decir, “una técnica muy adecuada para extraer información de un conjunto de datos sin imponer restricciones previas en forma de modelos estadísticos, al menos de manera explícita y, por ello, puede llegar a ser muy útil como una herramienta de elaboración de hipótesis acerca del problema considerado sin imponer patrones o teorías previamente establecidas” (Quintín, Cabero Morán, & De Paz Santana, 2008, pág. 298).

Ambos análisis están relacionados, ya que los componentes que se construyen corresponden a los indicadores de pertenencia al clúster y se utilizan para representar visualmente los grupos creados (Adolfsson, Ackerman, & Brownstein, 2018).

## **Datos y metodología**

### **Tabla de Datos**

La tabla de datos (ver **Anexo 1**) fue construida con información extraída del Programa de las Naciones Unidas para el Desarrollo del año 2020 (PNUD, 2020). Es de corte transversal ya que está compuesta por 15 variables cuantitativas y 117 unidades de observación, las cuales son países (Anexo 2). Las observaciones con celdas vacías fueron eliminadas del análisis.

En el siguiente cuadro se observan las variables escogidas, las utilizadas identificaciones utilizadas para cada una de ellas, su año de medición y definición.

**Tabla 1**

*Tabla de datos*

Dim	Identif.	Variable	Año	Descripción
Salud	EVN	Esperanza de Vida al Nacer	2019	Número de años que podría esperar vivir un recién nacido si los patrones predominantes de tasas de mortalidad específicas por edad en el momento del nacimiento permanecen iguales durante toda la vida del bebé.
Educación	AEE	Años de Escolaridad Esperados	2019	Número de años de escolarización que un niño en edad de ingresar a la escuela puede esperar recibir si los patrones predominantes de tasas de matriculación específicas por edad persisten a lo largo de la vida del niño.
Educación	AME	Años Medios de Escolaridad	2019	Número promedio de años de educación recibidos por personas de 25 años o más, convertidos a partir de los niveles de logro educativo utilizando las duraciones oficiales de cada nivel.
Economía	CAP	Formación Bruta de Capital	2019	Desembolsos por adiciones al activo fijo de la economía más variaciones netas de existencias como % del PBI. La formación



					bruta de capital se conocía anteriormente como inversión interna bruta.
Economía	DES	Tasa de Desempleo Total	2019		Porcentaje de la población activa de 15 años o más que no tiene un empleo remunerado o autónomo, pero que está disponible para trabajar y ha tomado medidas para buscar un empleo remunerado o por cuenta propia.
Economía	IPC	Índice de Precios al Consumidor	2019		Índice que refleja los cambios en el costo para el consumidor promedio de adquirir una canasta de bienes y servicios que pueden fijarse o cambiarse en intervalos específicos, como anualmente (2010=100).
Economía	PBIC	Producto Bruto Interno Per Cápita	2019		PIB en un período particular dividido por la población total en el mismo período (2017 PPC \$).
Desigualdad	CDH	Coeficiente de Desigualdad Humana	2019		Desigualdad media en las tres dimensiones básicas del desarrollo humano, economía, educación y salud
Desigualdad	GINI	Coeficiente de GINI			Medida de la desviación de la distribución del ingreso entre individuos u hogares en un país de una distribución perfectamente equitativa. Un valor de 0 representa igualdad absoluta, un valor de 100 desigualdad absoluta.

Desigualdad	IDG	Índice de Desigualdad de Género	de 2019	Una medida compuesta que refleja la desigualdad en el rendimiento entre mujeres y hombres en tres dimensiones: salud reproductiva, empoderamiento y mercado laboral.
Sociedad	PDA	Profundidad del Déficit de Alimentos (Adecuación Media del Suministro de Energía Alimentaria)	del 2019	Adecuación promedio del suministro de energía alimentaria: Suministro promedio de calorías para el consumo de alimentos, expresado como un porcentaje del requerimiento promedio de energía dietética, que es la cantidad de energía dietética que se necesitaría para garantizar que, si se distribuye adecuadamente, se elimine el hambre.
Sociedad	POB	Población Total	Millones	Población de hecho en un país, área o regiones al 1 de julio.
Gobierno	GCF	Gasto de Consumo Final del Gobierno General	2019	Todos los gastos corrientes del gobierno para la compra de bienes y servicios (incluida la compensación de los empleados y la mayoría de los gastos en defensa y seguridad nacional, pero excluidos los gastos militares del gobierno que forman parte de la formación de capital del gobierno), expresados como un porcentaje del PIB.

Medioambiente	CO2C	Emisiones de 2018	Emisiones totales de dióxido de carbono
		Dióxido de	producidas como consecuencia de las
		Carbono Per Cápita	actividades humanas (uso de carbón, petróleo y gas para procesos de combustión e industriales, quema de gas y fabricación de cemento), dividido por la población a mitad de año. Los valores son emisiones territoriales, lo que significa que las emisiones se atribuyen al país en el que se producen físicamente.

---

*Fuente. Elaboración propia con datos extraídos del PNUD (Programa de las Naciones Unidas para el Desarrollo)*

Para la realización de los análisis se realizará la estandarización de los datos, buscando lograr variables comparables y evitando que nos provean información errónea.

## **Metodología**

La consecución de los objetivos exige la utilización de un programa o software y llevar a cabo diferentes pasos en cada análisis multivariante.

Se realizarán los análisis mediante el programa RStudio, un software estadístico gratuito y de código abierto para ciencia de datos, investigación científica y comunicación técnica.

## ***Análisis de Componentes Principales***

Luego de construir la tabla de datos y estandarizarla realizamos el análisis de componentes principales:

- Construcción y análisis de la matriz de correlación y test KMO: a través de estos métodos determinamos si nuestros datos pueden analizarse a través del método de Análisis de Componentes Principales, analizando si existe relación entre las variables en análisis.
- Selección de componentes: luego de aplicar ACP a los datos normalizados, seleccionamos la cantidad de componente significativos para nuestro análisis a través del gráfico de sedimentación que ilustra el tamaño de los auto valores, es decir, el porcentaje de variabilidad que cada componente captura.
- Matriz de rotación: obtenemos los vectores que nos permiten a través de los coeficientes de relación entre las variables y los componentes seleccionados, interpretar la composición de cada uno.
- Interpretación de los componentes y caracterización de individuos: para cada componente definimos la asociación de variables en su semieje positivo y negativo, lo que nos permite ubicar a nuestras unidades de observación opuestas en cada uno de ellos y construir su ecuación. Aplicaremos la información obtenida caracterizando dos países.
- Construcción de las nuevas variables: terminado el análisis de ACP podemos construir nuestras nuevas variables, que suplantando a las originales. De esta manera reducimos la dimensionalidad de nuestros datos para facilitar y agilizar su análisis.

### ***Análisis de Cluster***

Una vez finalizado el ACP realizaremos un estudio de cluster para agrupar las unidades de observación:

- Análisis de tendencia de agrupación: construimos la matriz de distancia y realizamos el test de Hopkins para observar si existe tendencia entre los datos analizados, es decir, si estos son agrupables.
- Determinación del número de clusters: realizamos el método de la “rodilla”, la silueta y la estadística de brecha para determinar la cantidad de particiones óptimas, que luego procedemos a graficar.
- Método k-medias e interpretación de resultados: aplicamos el algoritmo k-medias, que nos provee la información necesaria para caracterizar cada cluster y definir así los atributos de las unidades de observación que los componen. Por último, representamos gráficamente los resultados obtenidos.

## Resultados

### Tabla de Datos

La aplicación de los análisis requiere de una tabla de datos bien especificada y variables definidas para lograr un tratamiento adecuado de la información.

```
datos=read.csv("../TP final/Trabajo R.csv",header=TRUE,sep=";",dec=",")
rownames(datos)<-datos$OBS.VAR
datos$OBS.VAR<-NULL
```

Luego de cargar nuestra tabla podemos obtener la cantidad de unidades de observación y variables que esta contiene.

```
dim(datos)
```

```
## [1] 117 15
```

```
head(datos, 117)
```

## *Análisis de la Varianza*

Para determinar si realizaremos nuestros análisis con variables estandarizadas comparamos la varianza de cada una de ellas.

Como podemos observar las varianzas de las 15 variables tienen magnitudes muy distantes y se miden en diferentes escalas, lo que nos puede guiar a obtener resultados sesgados. Para evitar este resultado es recomendable estandarizar los valores, centrando (restando la media) y escalando (dividiendo por la desviación estándar) cada una de las variables para reducir su variabilidad y convertirlas en comparables.

```
round(apply(datos,2,var),2)
```

##	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC
##	59.06	9.14	10.02	59.79	28.98	360314530.05	15749.99	426419098.59
##	CDH	GINI	IDG	PDA	POB	GCF	CO2C	
##	90.85	60.69	0.04	241.12	34543.97	27.54	17.29	

## **Análisis de Componentes Principales**

### *Construcción y análisis de la matriz de correlación y test KMO*

**Matriz de Correlación.** La matriz de correlación nos permitirá evaluar la fuerza y dirección de la relación entre las variables tipificadas seleccionadas. Un valor de correlación elevado y positivo nos indica que los elementos miden la misma característica, caso contrario si los elementos no están altamente correlacionados, donde nuestras variables pudieran medir diferentes atributos o no estar claramente definidas.

La matriz será de doble entrada, y mostrará una lista multivariable horizontalmente y la misma lista verticalmente, en su intersección encontraremos los coeficientes de correlación, “r”. Éstos pueden tomar un valor entre -1 y 1, donde valores cercanos a cero muestran que no existe

relación lineal, cercanos a 1 que las dos variables están asociadas en sentido directo (elevada asociación positiva) y cercanos a -1 en sentido inverso (elevada asociación negativa).

En la diagonal encontramos valores iguales a 1, ya que representan la relación entre una variable y sí misma.

**Coefficiente de Correlación de Pearson.** Calcularemos la matriz mediante el método de Pearson, cuyos coeficientes de correlación se definen como la covarianza existente entre dos variables tipificadas y se obtienen sumando los productos de cada pareja de datos y dividiendo el resultado por el producto del número de datos por la desviación estándar de cada variable.

```
matriz<-rcorr(as.matrix(datos))
matrizr<-round(matriz$r,2)
matrizr
```

**Tabla 2**

*Matriz de correlación*

	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC	CDH	GINI	IDG	PDA	POB	GCF	CO2C
EVN	1.00	0.82	0.80	-0.05	-0.11	0.78	-0.21	0.75	-0.85	-0.44	-0.86	0.64	0.01	0.28	0.61
AEE	0.82	1.00	0.83	-0.10	-0.01	0.79	-0.25	0.75	-0.81	-0.35	-0.84	0.62	-0.03	0.43	0.67
AME	0.80	0.83	1.00	-0.11	0.01	0.79	-0.24	0.75	-0.91	-0.42	-0.87	0.48	-0.07	0.39	0.71
CAP	-0.05	-0.10	-0.11	1.00	-0.18	-0.08	-0.03	-0.07	0.03	-0.08	0.08	-0.01	0.21	-0.16	-0.03
DES	-0.11	-0.01	0.01	-0.18	1.00	-0.16	0.18	-0.16	0.10	0.23	0.03	-0.02	-0.09	0.24	-0.08
INB	0.78	0.79	0.79	-0.08	-0.16	1.00	-0.24	0.98	-0.76	-0.46	-0.84	0.62	-0.05	0.41	0.74
IPC	-0.21	-0.25	-0.24	-0.03	0.18	-0.24	1.00	-0.22	0.24	0.05	0.23	-0.11	0.02	-0.26	-0.14
PBIC	0.75	0.75	0.75	-0.07	-0.16	0.98	-0.22	1.00	-0.73	-0.43	-0.80	0.59	-0.06	0.37	0.75
CDH	-0.85	-0.81	-0.91	0.03	0.10	-0.76	0.24	-0.73	1.00	0.61	0.89	-0.53	0.05	-0.38	-0.67
GINI	-0.44	-0.35	-0.42	-0.08	0.23	-0.46	0.05	-0.43	0.61	1.00	0.51	-0.36	0.04	-0.15	-0.40
IDG	-0.86	-0.84	-0.87	0.08	0.03	-0.84	0.23	-0.80	0.89	0.51	1.00	-0.57	0.03	-0.43	-0.68
PDA	0.64	0.62	0.48	-0.01	-0.02	0.62	-0.11	0.59	-0.53	-0.36	-0.57	1.00	0.02	0.16	0.52
POB	0.01	-0.03	-0.07	0.21	-0.09	-0.05	0.02	-0.06	0.05	0.04	0.03	0.02	1.00	-0.10	0.05
GCF	0.28	0.43	0.39	-0.16	0.24	0.41	-0.26	0.37	-0.38	-0.15	-0.43	0.16	-0.10	1.00	0.34
CO2C	0.61	0.67	0.71	-0.03	-0.08	0.74	-0.14	0.75	-0.67	-0.40	-0.68	0.52	0.05	0.34	1.00

**Contraste de Hipótesis Nula de Aleatoriedad.** Los p-valores muestran la significancia de los niveles de correlación, y nos permiten analizar la aleatoriedad de la correlación. Considerando un intervalo de confianza del 95% realizamos un contraste normal de la siguiente hipótesis nula, ‘Las correlaciones son aleatorias’.

```
matrizp<-round(matriz$P,2)
matrizp
```

**Tabla 3**

*P-valores de la matriz de correlación*

	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC	CDH	GINI	IDG	PDA	POB	GCF	CO2C
EVN	NA	0.00	0.00	0.63	0.25	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00
AEE	0.00	NA	0.00	0.30	0.90	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.73	0.00	0.00
AME	0.00	0.00	NA	0.23	0.95	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.45	0.00	0.00
CAP	0.63	0.30	0.23	NA	0.05	0.39	0.73	0.46	0.71	0.39	0.37	0.95	0.03	0.09	0.72
DES	0.25	0.90	0.95	0.05	NA	0.09	0.05	0.10	0.30	0.01	0.78	0.85	0.33	0.01	0.41
INB	0.00	0.00	0.00	0.39	0.09	NA	0.01	0.00	0.00	0.00	0.00	0.00	0.56	0.00	0.00
IPC	0.03	0.01	0.01	0.73	0.05	0.01	NA	0.02	0.01	0.61	0.01	0.24	0.83	0.00	0.14
PBIC	0.00	0.00	0.00	0.46	0.10	0.00	0.02	NA	0.00	0.00	0.00	0.00	0.52	0.00	0.00
CDH	0.00	0.00	0.00	0.71	0.30	0.00	0.01	0.00	NA	0.00	0.00	0.00	0.60	0.00	0.00
GINI	0.00	0.00	0.00	0.39	0.01	0.00	0.61	0.00	0.00	NA	0.00	0.00	0.68	0.12	0.00
IDG	0.00	0.00	0.00	0.37	0.78	0.00	0.01	0.00	0.00	0.00	NA	0.00	0.73	0.00	0.00
PDA	0.00	0.00	0.00	0.95	0.85	0.00	0.24	0.00	0.00	0.00	0.00	NA	0.86	0.09	0.00
POB	0.88	0.73	0.45	0.03	0.33	0.56	0.83	0.52	0.60	0.68	0.73	0.86	NA	0.30	0.57
GCF	0.00	0.00	0.00	0.09	0.01	0.00	0.00	0.00	0.00	0.12	0.00	0.09	0.30	NA	0.00
CO2C	0.00	0.00	0.00	0.72	0.41	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.57	0.00	NA

Como podemos observar los p-valores en su mayoría muestran aceptación de la hipótesis nula, como la intersección de las variables Esperanza de Vida al Nacer (EVN) y Coeficiente de Desigualdad Humana (CDH) cuya correlación es igual a  $-0,85$ , y su valor test es de  $0,00$ . Todos aquellos p-valores menores a  $0,05$  son variables que están correlacionadas aleatoriamente.

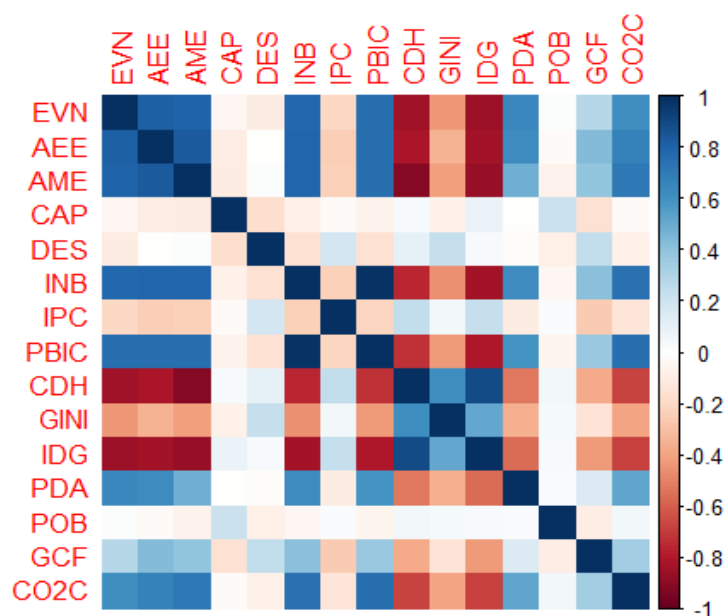


**Correlograma.** A continuación, podemos observar un diagrama de correlación que a través de diferentes intensidades de color rojo y azul, ilustra el valor los coeficientes de correlación.

```
corrplot(matrizr,method = "color",type=c("full"))
```

**Figura 1**

*Correlograma*



Podemos observar que aquellas variables representadas con color azul oscuro presentan fuertes asociaciones positivas, y aquellas con rojo oscuro negativas, a su vez, a medida que disminuye la intensidad del color, disminuye la fuerza de correlación entre las variables.

La asociación entre la variable Esperanza de Vida al Nacer (EPV) y Coeficiente de Desigualdad Humana (CDH) es elevada y negativa, es decir, que a mayor esperanza de vida menor es el coeficiente de desigualdad, y viceversa, a mayor desigualdad, menor esperanza de vida.

Otro ejemplo muy ilustrativo es la relación entre Emisiones de Dióxido de Carbono Per Cápita (CO2C) y Producto Bruto Interno Per Cápita (PBIC), que como podemos observar están

relacionadas de forma positiva, es decir, a mayor Producto Bruto Interno mayores son las emisiones de Dióxido de Carbono, en términos per cápita.

Por otro lado, las variables Formación Bruta de Capital (CAP), Tasa de Desempleo Total (DES) y Población Total (POB) muestran bajos niveles de correlación.

**Método KMO.** Este método relaciona los coeficientes de correlación, cuanto más cercano a 1 es el valor obtenido de KMO, la relación entre las variables será más elevada. La toma de decisión de realizar el ACP dependerá de los siguientes valores de calidad del KMO:

- $KMO \geq 0.9$ , el test es muy bueno
- $KMO \geq 0.8$ , el test es notable
- $KMO \geq 0.7$ , el test es mediano
- $KMO \geq 0.6$ , el test es bajo
- $KMO < 0.5$ , el test es muy bajo

Si se acepta la hipótesis nula, “se puede aplicar el análisis factorial”.

```
KMO(datos)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = datos)
## Overall MSA = 0.85
## MSA for each item =
## EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG PDA POB GCF CO2C
## 0.92 0.95 0.84 0.68 0.33 0.81 0.76 0.82 0.82 0.74 0.95 0.88 0.45 0.71 0.94
```

El valor de  $MSA = 0.85$ , por lo cual nuestro resultado no sólo es mayor a 0,5 sino que es notable y nos ordena realizar el análisis de componentes principales.

### *Selección de Componentes*

Aplicamos el ACP a los datos normalizados. La función ‘summary’ nos provee un resumen del análisis destacando la importancia de cada componente en el conjunto de datos, mediante el porcentaje de variabilidad que cada una de ellas captura.

```
acp<-prcomp(datos,center=TRUE,scale=TRUE)
summary (acp)
```

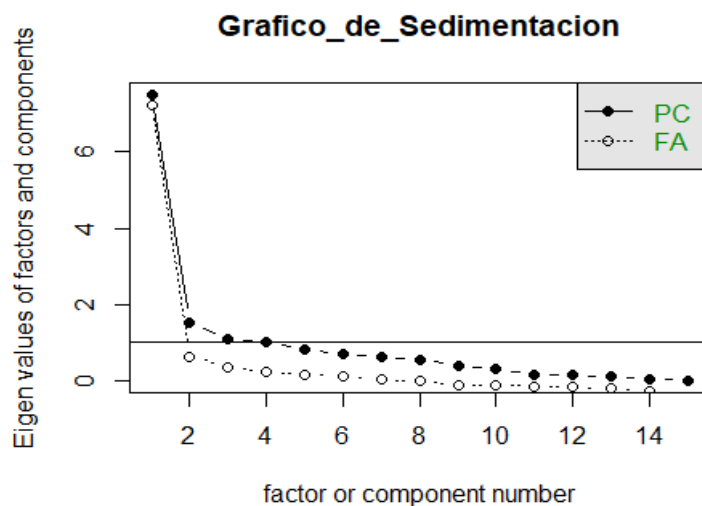
```
## Importance of components:
##
## Standard deviation      2.7403 1.2390 1.04251 1.01053 0.90212 0.84280 0.77681 0.74829 0.63701 0.55009 0.39371
## Proportion of Variance 0.5006 0.1023 0.07245 0.06808 0.05425 0.04735 0.04023 0.03733 0.02705 0.02017 0.01033
## Cumulative Proportion 0.5006 0.6029 0.67541 0.74349 0.79774 0.84510 0.88532 0.92265 0.94971 0.96988 0.98021
##
## Standard deviation      0.36964 0.31998 0.21103 0.11510
## Proportion of Variance 0.00911 0.00683 0.00297 0.00088
## Cumulative Proportion 0.98932 0.99615 0.99912 1.00000
```

**Selección de Componentes Significativos.** El gráfico de sedimentación representa gráficamente el tamaño de los auto valores, a mayor inclinación mayor es la variabilidad explicada por los componentes, como podemos el punto de inflexión en el cual los componentes comienzan a formar una planicie, es en el componen ente 3 como podemos observar a continuación.

```
scree(datos,main ="Grafico_de_Sedimentacion")
```

## Figura 2

*Gráfico de sedimentación*



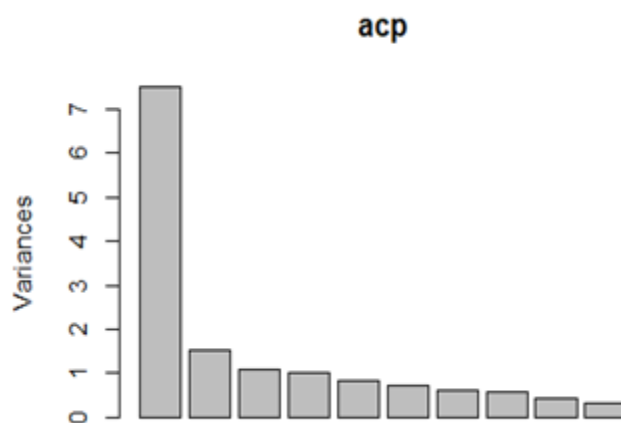
Finalmente, seleccionaremos los tres primeros componentes, los cuales tienen una variabilidad de 50,1%, 10,23% y 7,25% respectivamente. Acumulando estas varianzas explicamos aproximadamente el 68% de la variabilidad.

En el siguiente gráfico de barras se observa la importancia de cada componente. Las alturas son las varianzas de las correspondientes componentes.

```
plot(acp)
```

**Figura 3**

*Gráfico de barras de varianza capturada por cada componente*

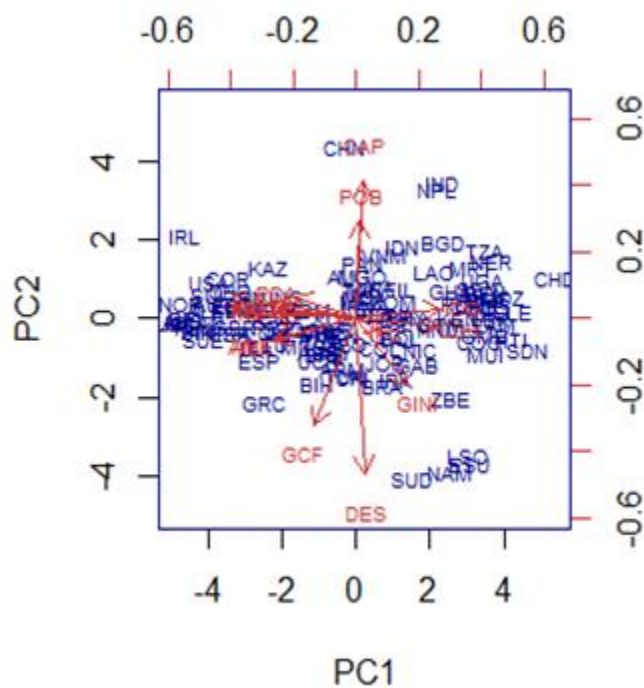


En el siguiente diagrama podemos observar la ubicación en de las dos primeras componentes de las variables (rojo) y de las unidades de observación (azul).

```
biplot(acp,scale=0,cex=0.6,col=c("blue4","brown3"))
```

**Figura 4**

*Representación de las variables y las unidades de observación en las dos primeras componentes*



**Matriz de Rotación**

**Vectores Propios.** Obtenemos la matriz de rotación que muestra los coeficientes de combinaciones lineales de las variables con los componentes, llamados vectores propios.

```
round(acp$rot[,1:3],2)
```

**Tabla 4**

*Valores propios de los componentes principales seleccionados*

	PC1	PC2	PC3
EVN	-0.33	0.06	-0.08
AEE	-0.33	-0.08	0.00
AME	-0.33	-0.08	0.00
CAP	0.03	0.52	0.04
DES	0.04	-0.58	-0.31
INB	-0.34	0.02	0.01
IPC	0.10	-0.04	-0.83
PBIC	-0.33	0.03	0.01
CDH	0.34	-0.03	0.01
GINI	0.20	-0.25	0.13
IDG	0.34	0.04	0.03
PDA	-0.25	0.08	-0.24
POB	0.02	0.37	-0.16
GCF	-0.17	-0.40	0.31
CO2C	-0.29	0.04	-0.10

### *Interpretación de los Componentes y Caracterización de Individuos*

**Interpretación de cada Componente.** A continuación, describiremos analizando la anterior matriz las características de cada componente:

- Primer componente: tiene asociaciones altas en el semieje negativo con las variables el Ingreso Nacional Bruto y Producto Bruto Interno Per Cápita, las correspondientes a las dimensiones de Salud, Educación, Medioambiente, y la

Profundidad del Déficit de Alimentos; y en el semieje positivo las variables de desigualdad.

Esta componente opondrá a los países con alto nivel de ingreso nacional bruto y producto bruto interno per cápita, elevada esperanza de vida, cantidad de años de educación y, niveles de emisiones de CO2 per cápita y un alto porcentaje del requerimiento promedio de energía dietética que se necesitaría para garantizar que se elimine el hambre; con los países con un elevado coeficiente de desigualdad humana, que dimensiona a la desigualdad económica, educacional y en salud, e índice de desigualdad de género.

$$\begin{aligned} Comp1 = & -0,34(INB) - 0,33(PBIC) - 0,33(EVN) - 0,33(AEE) - 0,33(AME) \\ & - 0,29(CO2C) - 0,25(PDA) + 0,34(CDH) + \dots + 0,34(IDG) \end{aligned}$$

- Segundo componente: tiene asociaciones altas en el semieje negativo con las variables Tasa de Desempleo Total, Gasto de Consumo Final del Gobierno General y Coeficiente de GINI; y en el semieje positivo las variables Formación Bruta de Capital y Población Total.
- Esta componente opondrá a los países con elevado desempleo, gasto corriente público en bienes y servicios y un coeficiente de GINI elevado, lo cual expresa que la distribución del ingreso entre individuos u hogares en el país es más cercana a la desigualdad absoluta que a la equidad; con los países con elevada inversión interna bruta y población.

$$Comp2 = -0,58(DES) - 0,40(GCF) - 0,25(GINI) + \dots + 0,52(CAP) + 0,37(POB)$$

- Tercer componente: tiene asociaciones altas en el semieje negativo con las variables Índice de Precios al Consumidor, Tasa de Desempleo Total y Profundidad del Déficit de Alimentos; y en el semieje positivo la variable Gasto de Consumo Final del Gobierno General.

Esta componente opodrá a los países con principalmente prominente inflación medida a través del índice de precios al consumidor, y con menor peso en su elevada tasa de desempleo y profundidad del déficit de alimentos; con los países con elevado gasto corriente público en bienes y servicios.

$$Comp3 = -0,83(IPC) - 0,31(DES) - 0,24(PDA) + \dots + 0,31(GCF)$$

Podemos determinar la ubicación de cada país en los componentes analizados y describir así las características de cada uno, sólo analizando 3 variables independientes (ver **Anexo 2**).

```
componentes<-round(acp$x[,1:3],2)
head(componentes,117)
```

## Caracterización de Individuos.

### *Argentina*

```
Argentina<-head(componentes[4,])
Argentina
```

```
##   PC1   PC2   PC3
## -0.98 -0.85 -0.89
```

Este país se ubica en el semieje negativo en los primeros tres componentes y en el positivo en el último, podemos afirmar que presenta:



- Elevado nivel de ingreso nacional bruto y producto bruto interno per cápita, elevada esperanza de vida, cantidad de años de educación, niveles de emisiones de CO2 per cápita y un profundo déficit de alimento.
- Elevado desempleo, gasto corriente público en bienes y servicios y un coeficiente de GINI elevado, lo cual expresa que la distribución del ingreso es cercana a la desigualdad absoluta.
- Prominente inflación medida a través del índice de precios al consumidor, elevada tasa de desempleo y profundidad del déficit de alimentos.

### ***Tanzania***

```
Tanzania<-head(componentes[109,])
Tanzania
```

```
## PC1 PC2 PC3
## 3.45 1.74 0.18
```

Este país se ubica en el semieje positivo en los tres componentes, podemos afirmar que presenta:

- Un elevado coeficiente de desigualdad humana, que dimensiona a la desigualdad económica, educacional y en salud, e índice de desigualdad de género.
- Elevada inversión interna bruta y población.
- Elevado gasto corriente público en bienes y servicios, que dado su cercanía al centro de gravedad (0,18), podemos afirmar que es muy cercano al individuo promedio descrito por esta componente.

### ***Construcción de las Nuevas Variables***

Construimos las nuevas variables creadas, sumando el producto de todas las filas de la primera columna de la matriz de rotación, por los datos originales.

```
pc1<-apply(acp$rotation[,1]*datos,1,sum)
pc2<-apply(acp$rotation[,2]*datos,1,sum)
pc3<-apply(acp$rotation[,3]*datos,1,sum)
datos$pc1<-pc1
datos$pc2<-pc2
datos$pc3<-pc3
datos[,1:15]<-NULL
```

Finalmente obtenemos una nueva tabla de datos que explica el 68% de la información original, y contiene sólo 3 variables independientes, que nos permiten caracterizar las 117 unidades de observación (ver **Anexo 3**).

```
round(head(datos,117),2)
```

## Análisis de Cluster

La tabla de datos será utilizada de forma estandarizada por razones mencionadas en el análisis previo.

### *Análisis de Tendencia de Agrupación*

**Matriz de Distancia.** La similitud o diferencia entre las diferentes observaciones se mide a través de la distancia entre estas, entendiendo mayor cercanía como mayor similitud.

Elegimos una distancia y no una medida de correlación como medida de asociación, por lo tanto, los grupos que creados se compondrán de individuos similares, es decir, de cercanos. (Sánchez López, Cruz-Gutiérrez, Posada-Zamora, Torrijos M., & Osorio Lama, 2016).

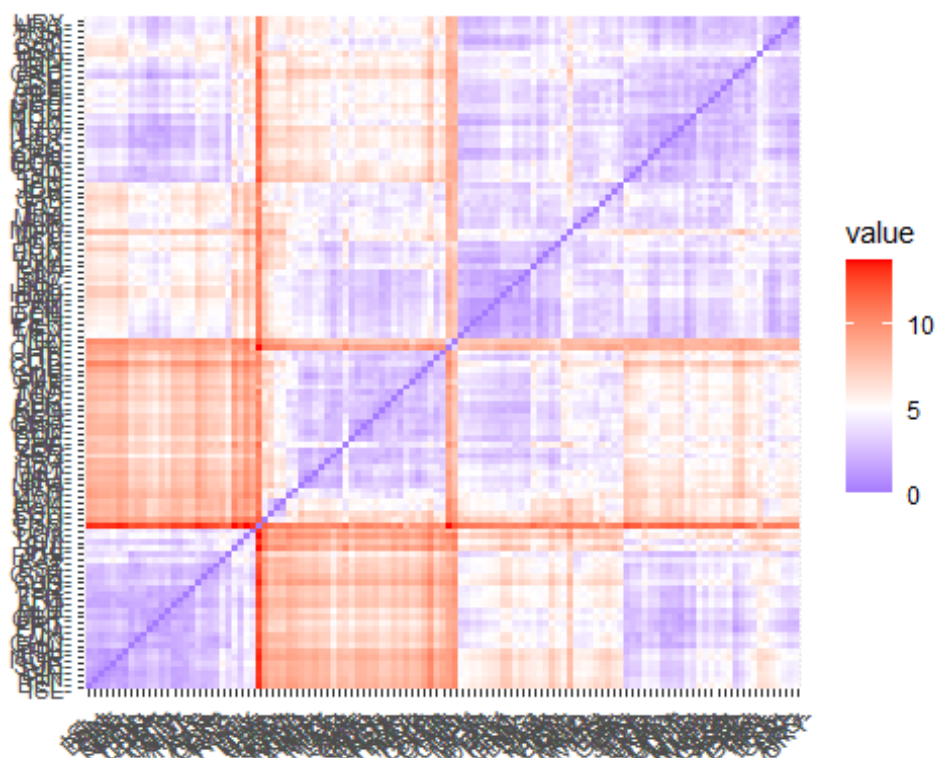
El método utilizado será la distancia euclídea, cuya principal ventaja es dar mayor influencia a aquellas observaciones que se encuentran más distantes.

El siguiente gráfico representa en un gradiente de colores, el grado de similitud, distancia, entre observaciones.

```
distancia<-get_dist(datos2,method ="euclidean")
fviz_dist(distancia,gradient=list(low="blue",mid="white",high="red"))
```

**Figura 5**

*Matriz de distancia*



Se pueden observar tendencias entre las unidades de observación, áreas de fuerte similitud, identificadas por el color rojo, y áreas desiguales por el color azul. Podemos afirmar que nuestros datos se pueden clusterizar.

**Test Estadístico de Hopkins.** Otra forma de evaluar la tendencia de agrupamiento de un conjunto de datos, es el test estadístico de Hopkins. Este método que prueba la aleatoriedad espacial de los datos, se calcula como “la distancia media del vecino más cercano en el conjunto de datos aleatorios dividida por la suma de la media del vecino más cercano” (Kassambara, 2017).

La  $H_0$  de este estadístico es que “el conjunto de datos está distribuido uniformemente (es decir, no contiene agrupaciones significativas)”. Buscamos rechazar esta hipótesis, determinando que nuestros datos son agrupables.

```
hopkins(datos2, n=nrow(datos2)-1)
```

```
## $H
## [1] 0.1908029
```

La  $H_0$  se rechaza para valores de estadístico menores a 0,5. Como podemos observar nuestros datos son aptos para agrupar, ya que el valor  $H = 0,19$ , valor cercano a 0.

### ***Determinación del Número de Clusters***

El usuario debe especificar la cantidad de cluster que se utilizarán, en consecuencia realizaremos diferentes pruebas para determinar el número óptimo de estos.

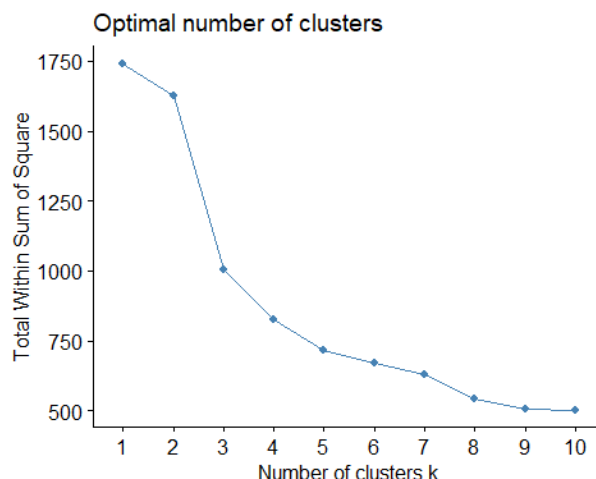
**Método de la “Rodilla”.** Este método busca minimizar la variación total dentro de cada conglomerado -la suma total del cuadrado dentro de este- y la representa en un gráfico. Dicha variación mide que tan compacto es el conglomerado, y nos provee de información para seleccionar la cantidad de clusters de modo que estos sean lo más compactos posibles.

La suma de errores cuadráticos es la suma del error cuadrado para cada punto -el cuadrado de la distancia del punto de su representación, es decir, su centro de clúster previsto- de todos los puntos.

```
fviz_nbclust(datos2, kmeans, method = "wss")
```

## **Figura 6**

*Gráfico del método de la “rodilla”*



La varianza disminuye a medida que aumenta el número de Cluster, a partir de la curva (o “codo”) los grupos tienen un valor pequeño. Este método nos indica que el número óptimo de cluster 3, ya que es allí donde se ubica dicha curva (rodilla).

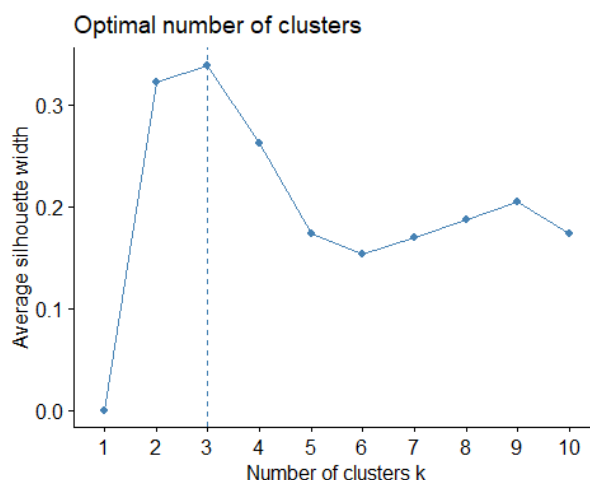
**Método de la Silueta.** Este método busca determinar cuan bien se encuentra cada individuo dentro de su cluster, un ancho de silueta alto indica una buena agrupación.

Seleccionaremos aquél número de conglomerados que maximice la silueta promedio.

```
fviz_nbclust(datos2, kmeans, method = "silhouette")
```

## Figura 7

*Gráfico del método de la silueta*



Como se observa, la cantidad de conglomerados que nos provee la máxima silueta promedio es 3, y como segundo número 2.

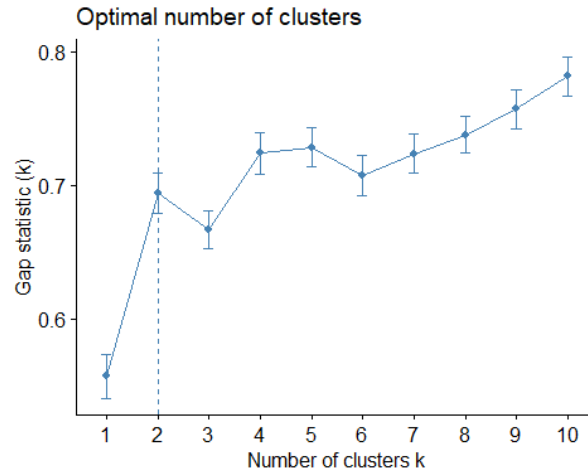
**Método de Estadística de Brecha.** Este análisis compara “la variación intragrupo total para diferente cantidad de clusters con sus valores esperados bajo una distribución de referencia nula de los datos (es decir, una distribución sin agrupamiento obvio)” (Kassambara, 2017).

El siguiente gráfico nos permite observar estadística de brecha y el error estándar. La cantidad de conglomerados que deben seleccionarse será aquel que maximice la estadística de brecha, es decir, que nos asegura que la estructura de agrupamiento está muy lejos a partir de la distribución uniforme aleatoria de puntos.

```
fviz_nbclust(datos2, kmeans, method = "gap_stat")
```

**Figura 8**

*Gráfico del método de estadística de brecha*



Podemos observar que la cantidad de cluster que maximiza la brecha es igual a 2.

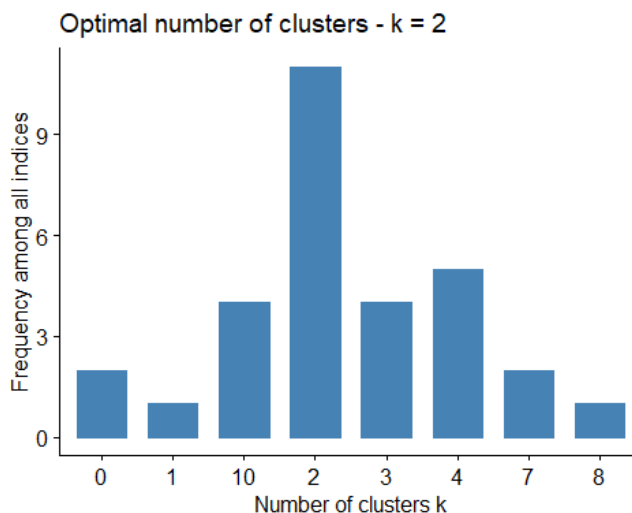
**Paquete de Pruebas Estadísticas de R.** Luego de realizar tres pruebas y no obtener resultados definitivos, realizaremos un paquete de 30 pruebas que provee RStudio para determinar cuál es número óptimo de cluster que debemos elegir.

```
resnumclust<-NbClust(datos2, distance = "euclidean", min.nc=2, max.nc=10, method = "kmeans", index = "alllong")
```

```
## *****
## * Among all indices:
## * 11 proposed 2 as the best number of clusters
## * 4 proposed 3 as the best number of clusters
## * 5 proposed 4 as the best number of clusters
## * 2 proposed 7 as the best number of clusters
## * 1 proposed 8 as the best number of clusters
## * 4 proposed 10 as the best number of clusters
##
##          ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is 2
```

**Figura 9**

*Gráfico de barras del número óptimo de clusters*



La mayoría de las pruebas nos recomiendan escoger 2 o 4 clusters.

### Representación Gráfica.

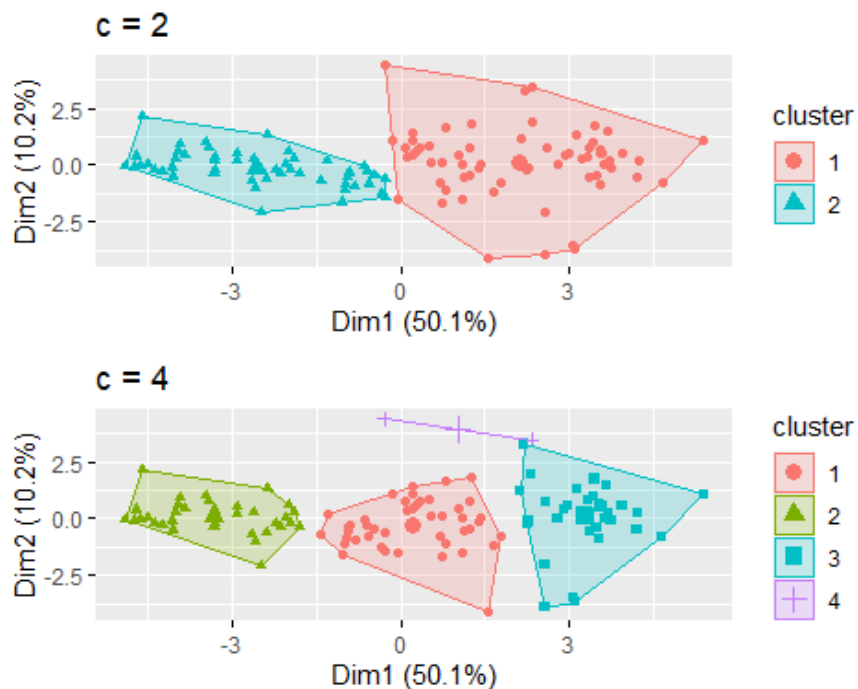
A continuación, se puede comparar el resultado de seleccionar cada opción mediante su representación gráfica.

```
cluster2<-kmeans(datos2,centers=2,nstart = 25)
cluster4<-kmeans(datos2,centers=4,nstart = 25)
c2 <- fviz_cluster(cluster2, geom = "point", data = datos2) + ggtitle("c = 2")
c4 <- fviz_cluster(cluster4, geom = "point", data = datos2) + ggtitle("c = 4")
grid.arrange(c2, c4, nrow = 2)
```

### Figura 10

*Representación gráfica de 2 y 4 clusters*





### *Método K-medias e Interpretación de Resultados*

**Método K-medias.** El cálculo de Cluster puede realizarse a través de dos tipos de algoritmos, jerárquicos y no jerárquicos o basados en particiones. Utilizaremos K-means, un método iterativo perteneciente al segundo grupo que “se basa en formar clusters de forma que la varianza interna de cada uno sea mínima” (Gómez Sánchez, 2018).

Este método asigna las N observaciones a los K clusters de modo que dentro de cada uno de ellos el promedio de las diferencias de cada individuo a la media del cluster, definido por los puntos del cluster, sea mínima **Fuente especificada no válida..**

Este algoritmo sigue la siguiente secuencia:

- 1 - Parte de un número de cluster que debe especificar el usuario
- 2 - Coloca las observaciones en cada uno de ellos
- 3 - Calcula los centroides de cada uno

4 - Reasigna cada observación al centroide más cercano

5 - Repite el paso 3

6 - Repite el paso 4 y 5 hasta que no sea posible mejorar los resultados, es decir, en el momento en que no haya más cambios de la ubicación de las observaciones entre los diferentes grupos para dos repeticiones sucesivas

Este método iterativo proporciona información sobre la media de cada variable de cada cluster, su composición, y la suma de cuadrados explicada por el análisis.

**Calculo de k Clusters.** Observamos la información provista por K-medias definiendo 2 y 4 clusters.

```
cluster2<-kmeans(datos2,centers=2,nstart = 25)
cluster2
```

```
## Within cluster sum of squares by cluster:
## [1] 332.2846 767.0367
## (between_SS / total_SS = 36.8 %)
```

```
cluster4<-kmeans(datos2,centers=4,nstart = 25)
cluster4
```

```
## Within cluster sum of squares by cluster:
## [1] 272.439685 6.235006 350.220498 177.775810
## (between_SS / total_SS = 53.6 %)
```

Como se mencionó anteriormente el criterio de k-medias es minimizar la dispersión dentro del grupo y maximizar la dispersión entre grupos. Este resultado se puede observar en el valor de (between\_SS/total\_SS), que mide la varianza total del conjunto de datos que se explica por la aglomeración.

Al seleccionar 4 grupos la varianza explicada aumenta considerablemente.

Finalmente, para realizar un análisis más exhaustivo de las observaciones determinaremos que la cantidad de cluster escogidos será de 4.

**Composición de cada Cluster.** Para analizar la caracterización de cada conglomerado analizaremos las variables más predominantes, luego compararemos los valores de sus medias y mencionaremos los países que forman parte de cada uno de ellos.

```
ClusterMediaScale<-aggregate(datos2,by=list(CLUSTER=cluster4$cluster),mean)
ClusterMedia<-aggregate(datos,by=list(CLUSTER=cluster4$cluster),mean)
```

```
Cluster<-cluster4$cluster
datos3<-datos
datos3$Clusters<-Cluster
```

A continuación, se observan los valores medios originales de cada variable, que luego serán comparados con la media de cada cluster.

```
round(apply(datos,2,mean),2)
```

##	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC	CDH	GINI	IDG	PDA
##	73.45	13.88	9.10	24.26	7.23	21081.65	157.25	21858.26	17.85	37.58	0.32	121.97
##	POB	GCF	CO2C									
##	57.19	16.02	4.18									

## Caracterización de cada Cluster.

### *Cluster 1.*

```
round(ClusterMediaScale[1,],2)
```

##	CLUSTER	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC	CDH	GINI	IDG	PDA	POB	GCF	CO2C
## 1	1	0.17	0.07	0.13	-0.04	0.4	-0.33	0.01	-0.32	-0.13	0.19	0.05	0.03	-0.1	-0.2	-0.23

Este cluster agrupará a los países principalmente por su tasa de desempleo y en menor medida por su coeficiente de GINI, su esperanza de vida y los años medios de escolaridad de sus habitantes.

```
round(ClusterMedia[1,],2)
```

```
## CLUSTER EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG PDA POB GCF CO2C
## 1 1 74.72 14.11 9.52 23.93 9.37 14793.12 159.02 15174.2 16.57 39.03 0.33 122.43 38.06 14.94 3.25
```

Los países del primer grupo disponen de una tasa de desempleo en promedio un 2% superior al resto de países, son más desiguales con respecto a la distribución salarial, y presentan valores similares a la media en las dimensiones de salud y educación.

Lo conforman los siguientes 44 países.

```
cluster4$size [1]
```

```
## [1] 44
```

```
Cluster[Cluster==1]
```

```
## ALB ALG ARG ARM BGR BIH BOL BRA CHL COL CRI DOM ECU EGI FIL GAB GEO IDN IRK IRN JOR KIR LKA MCN MDA MEX MGO
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## MNG MUC NIC PAN PER PRY ROM SLV SRB SUD TAI TAJ TUN TUR UCR URY VNM
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

## *Cluster 2.*

```
round(ClusterMediaScale[2,],2)
```

```
## CLUSTER EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG PDA POB GCF CO2C
## 2 2 -0.02 -0.26 -0.57 1.65 -0.44 -0.51 -0.04 -0.5 0.3 0.07 0.13 -0.22 7.23 -0.36 0.08
```

Este conglomerado estará definido por la formación bruta de capital y la población, esta última será solo significativa para este cluster.

```
round(ClusterMedia[2,],2)
```

```
## CLUSTER EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG PDA POB GCF CO2C
## 2 2 73.3 13.1 7.3 37 4.85 11369.35 152.75 11435.5 20.7 38.15 0.35 118.5 1400.1 14.15 4.5
```

Tanto China como India se agrupan por su considerable población y formación bruta de capital fijo. La primera es en promedio un 2.448% mayor y la segunda representa un 10% más del PBI.

Este cluster está compuesto por los dos países anteriormente nombrados.

```
cluster4$size [2]
```

```
## [1] 2
```

```
Cluster[Cluster==2]
```

```
## CHN IND
##      2   2
```

### *Cluster 3.*

```
round(ClusterMediaScale[3,],2)
```

```
## CLUSTER EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG PDA POB GCF CO2C
## 3      3 -1.25 -1.17 -1.25 0.13 -0.14 -0.92 0.33 -0.88 1.26 0.59 1.14 -0.84 -0.15 -0.38 -0.87
```

Se agrupan los países en el tercer cluster según sus niveles de desigualdad, tanto en ingresos, salud y educación como de género. Por otro lado, la variable índice de precios al consumidor se considera en menor medida.

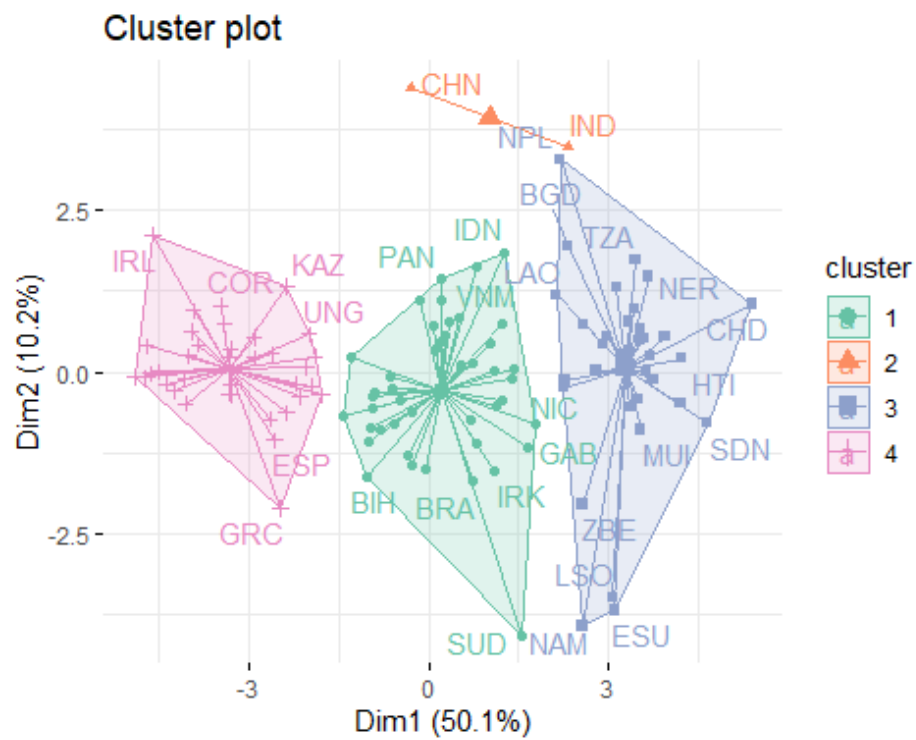
```
ClusterMedia[3,]
```

```
## CLUSTER EVN AEE AME CAP DES INB IPC PBIC CDH GINI IDG
## 3      3 63.84412 10.34412 5.144118 25.29118 6.482353 3599.006 198.7294 3695.068 29.86176 42.18824 0.5529412
##      PDA POB GCF CO2C
## 3 108.8824 28.64118 14.01765 0.5676471
```

El coeficiente de desigualdad humana en estos países es prominente, y alcanza aproximadamente los 30 puntos, por otro lado, el elevado coeficiente de GINI y de desigualdad de género explica una mayor desigualdad en la distribución del ingreso y en el rendimiento entre mujeres y hombres







Finalmente agrupamos a nuestras 117 observaciones en 4 conglomerados, minimizando la suma de cuadrados y maximizando la distancia entre sus vecinos.



### **Consideraciones Finales**

La creación de indicadores compuestos y el análisis de datos masivos requieren del uso de métodos multivariantes como los estudiados en este informe. La reducción de la dimensionalidad y la agrupación de individuos según sus semejanzas, facilita y agiliza el análisis de información y no presenta grandes dificultades en su aplicación.

Los resultados obtenidos nos permiten afirmar que los países del mundo agrupados en función de variables de las dimensiones de salud, educación, economía, desigualdad, sociedad y medio ambiente, se corresponden con las categorizaciones tradicionales de países desarrollados y subdesarrollados. El análisis de cluster es utilizado en amplias disciplinas, como investigaciones de cáncer, marketing y segmentación del mercado y planeamiento de ciudades (Kassambara, 2017).

Por otro lado, los componentes obtenidos pueden utilizarse para la construcción de modelos predictivos e indicadores compuestos. Éstos últimos son ampliamente empleados para la evaluación de desempeño de políticas públicas, universidades, estudios médicos, entre otros.

## Referencias

- Adolfsson, A., Ackerman, M., & Brownstein, N. C. (2018). *To cluster, not to cluster: an analysis of clusterability methods*. Florida.
- Baronio, A. M., & Vianco, A. M. (2014). *Análisis exploratorio: ACP interpretación de resultados*. Santa Rosa, La Pampa: Univerisdad de Villa María - Universidad de Río Cuarto.
- Colina, C. L., & Roldán, P. L. (1991). El análisis de componentes principales: aplicación al análisis de datos secundarios. *Revista de Sociología*, 31-63.
- Gómez Sánchez, J. A. (2018). *Análisis comparativo de diferentes métodos de agrupación para el tratamiento de datos de expresión genética*. Universidad Oberta de Cataluña.
- Jolliffe, I. T. (2002). *Principal component analysis*. New York, United State: Springer.
- Kassambara, A. (2017). *Practical guide to cluster analysis in R*. STHDA.
- OECD. (2008). *Handbook on constructing composite indicators: methodology and user guide*. Organisation for Econocic Co-Operation and Development Publishing.
- PNUD. (2020). *Informe sobre Desarrollo Humano 2020. La próxima frontera. El desarrollo Humano y el Antropoceno*. New York: Programa de Naciones Unidas para el Desarrollo (PNUD).
- Quintín, M. M., Cabero Morán, M. T., & De Paz Santana, Y. d. (2008). *Tratamiento estadístico de datos con SPSS. Prácticas resueltas y comentadas*. Madrid: Thomson.
- Rodríguez, P., Palomino, N., & Mondaca, J. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. Banco Interamericano de Desarrollo.
- Sánchez López, A., Cruz-Gutiérrez, V., Posada-Zamora, M. A., Torrijos M., M. T., & Osorio Lama, M. A. (2016). *Estudio del análisis de componentes principales en base de datos de calidad del aire*. México.

## Anexos

### Anexo 1

#### Tabla de datos

	EVN	AEE	AME	CAP	DES	INB	IPC	PBIC	CDH	GINI	IDG	PDA	POB	GCF	CO2C
ALB	78.6	14.7	10.1	25.1	12.3	13998.3	119.0	13961.8	10.9	33.2	0.2	138	2.9	11.5	1.6
ALG	76.9	14.6	8.0	44.3	11.7	11174.0	151.4	11349.7	19.7	27.6	0.4	145	43.1	18.6	3.7
ALM	81.3	17.0	14.2	21.6	3.0	55314.4	112.9	53815.4	7.9	31.9	0.1	138	83.5	20.3	9.1
ARG	76.7	17.7	10.9	18.2	9.8	21190.2	232.8	22033.9	13.2	41.4	0.3	134	44.8	15.3	4.4
ARM	75.1	13.1	11.3	17.4	17.0	13894.0	129.2	13653.7	9.7	34.4	0.2	127	3.0	12.5	1.9
AUS	83.4	22.0	12.7	23.3	5.3	48084.8	119.8	49756.3	7.9	34.4	0.1	134	25.2	18.9	16.9
AUT	81.5	16.1	12.5	25.4	4.7	56196.9	118.1	56352.4	6.9	29.7	0.1	145	9.0	19.3	7.7
BEL	81.6	19.8	12.1	25.3	5.6	52084.6	117.1	51707.8	7.7	27.4	0.0	148	11.5	23.2	8.7
BEN	61.8	12.6	3.8	25.6	2.2	3254.2	110.7	3287.3	36.9	47.8	0.6	124	11.8	10.3	0.6
BFA	61.6	9.3	1.6	26.0	6.3	2133.0	108.4	2189.6	29.5	35.3	0.6	122	20.3	14.4	0.2
BGD	72.6	11.6	6.2	31.6	4.2	4976.2	179.7	4753.7	23.7	32.4	0.5	109	163.0	6.3	0.5
BGR	75.1	14.4	11.4	19.5	4.3	23324.7	114.4	23091.3	11.3	40.4	0.2	113	7.0	17.5	6.3
BIH	77.4	13.8	9.8	21.3	18.4	14871.8	104.9	14894.8	14.2	33.0	0.1	130	3.3	20.0	6.5
BOL	71.5	14.2	9.0	19.9	3.5	8553.8	148.3	8724.5	23.7	42.2	0.4	106	11.5	18.1	2.0
BRA	75.9	15.4	8.0	15.1	12.1	14262.9	167.4	14651.6	24.4	53.9	0.4	132	211.0	20.3	2.2
CAN	82.4	16.2	13.4	22.7	5.6	48527.0	116.8	49031.4	8.4	33.8	0.1	139	37.4	21.2	15.3
CHD	54.2	7.3	2.5	21.4	1.9	1555.4	117.7	1579.6	37.4	43.3	0.7	95	15.9	3.6	0.1
CHI	81.0	15.2	12.2	19.1	7.3	38206.8	102.5	39544.7	9.1	31.4	0.1	103	1.2	16.4	6.3
CHL	80.2	16.4	10.6	22.8	7.1	23261.3	131.9	24226.2	15.9	44.4	0.2	122	19.0	14.6	4.6
CHN	76.9	14.0	8.1	43.8	4.3	16057.4	125.1	16116.7	15.7	38.5	0.2	128	1433.8	16.5	7.0
CHQ	79.4	16.8	12.7	26.3	1.9	38108.6	116.5	40314.2	4.4	24.9	0.1	128	10.7	20.2	9.9
CMF	57.8	10.0	5.3	21.0	3.3	5068.6	111.6	5238.3	35.3	41.5	0.6	123	25.7	9.6	0.3
CMR	59.3	12.1	6.3	24.2	3.4	3580.9	118.6	3652.7	33.4	46.6	0.6	120	25.9	11.3	0.3
COG	64.6	11.7	6.5	18.8	9.5	2879.2	124.7	3298.1	24.9	48.9	0.6	99	5.4	7.6	0.6
COL	77.3	14.4	8.5	22.3	9.7	14256.7	141.0	14730.9	21.6	50.4	0.4	131	50.3	15.5	2.0
COR	83.0	16.5	12.2	31.2	4.1	43043.7	115.2	42661.2	10.7	31.6	0.1	139	51.2	17.3	12.9
CRI	80.3	15.7	8.7	17.9	11.9	18486.4	128.8	19642.4	17.5	48.0	0.3	123	5.0	17.4	1.6
CRO	78.5	15.2	11.4	22.8	6.9	28069.8	109.8	28508.5	7.9	30.4	0.1	124	4.1	19.6	4.5

DIN	80.9	18.9	12.6	22.7	4.9	58661.9	110.3	57184.2	6.0	28.7	0.0	132	5.8	24.0	6.1
DOM	74.1	14.2	8.1	27.3	5.8	17591.1	135.5	18419.0	21.1	43.7	0.5	121	10.7	11.2	2.3
ECU	77.0	14.6	8.9	25.0	4.0	11044.0	124.1	11375.3	18.4	45.4	0.4	114	17.4	14.5	2.5
EGI	72.0	13.3	7.4	16.7	10.8	11466.1	264.4	11763.3	28.7	31.5	0.4	143	100.4	8.4	2.4
ESP	83.6	17.6	10.3	20.8	14.0	40974.5	111.0	40883.3	13.1	34.7	0.1	131	46.7	18.7	5.7
EST	78.8	16.0	13.1	28.1	5.1	36019.3	122.1	36710.4	6.9	30.4	0.1	127	1.3	19.9	14.8
ESU	60.2	11.8	6.9	13.1	22.1	7919.1	162.6	8688.1	29.0	54.6	0.6	104	1.1	22.8	1.1
EVN	81.3	17.6	12.7	20.7	4.2	38079.5	111.1	38688.8	4.6	24.2	0.1	129	2.1	18.4	6.9
EVQ	77.5	14.5	12.7	23.3	5.6	32113.0	115.3	32792.6	6.1	25.2	0.2	108	5.5	19.6	6.6
FIL	71.2	13.1	9.4	26.2	2.2	9777.8	129.6	8908.2	17.8	44.4	0.4	116	108.1	12.5	1.3
FIN	81.9	19.4	12.8	24.0	6.6	48511.4	112.3	48621.1	5.3	27.4	0.0	132	5.5	23.1	8.5
FRA	82.7	15.6	11.5	24.2	8.4	47172.5	110.0	46183.5	8.9	31.6	0.0	142	65.1	23.1	5.2
GAB	66.5	13.0	8.7	22.4	20.0	13929.8	122.6	14869.8	22.5	38.0	0.5	115	2.2	9.1	2.5
GBR	81.3	17.5	13.2	17.4	3.9	46070.6	119.6	46699.3	7.9	34.8	0.1	137	67.5	18.9	5.6
GEO	73.8	15.3	13.1	26.8	14.4	14428.8	133.6	15014.3	11.5	36.4	0.3	114	4.0	12.8	2.6
GHA	64.1	11.5	7.3	26.4	4.3	5268.9	268.4	5412.6	27.8	43.5	0.5	132	30.4	8.9	0.6
GMB	62.1	9.9	3.9	18.5	9.1	2167.9	172.7	2206.9	31.2	35.9	0.6	114	2.3	13.4	0.3
GRC	82.2	17.9	10.6	12.5	17.2	30154.6	101.9	30314.6	10.8	34.4	0.1	133	10.5	19.5	7.0
GTM	74.3	10.8	6.6	14.5	2.5	8494.1	142.9	8637.6	26.9	48.3	0.5	116	17.6	11.1	1.1
HND	75.3	10.1	6.6	22.3	5.4	5308.5	150.3	5728.0	24.8	52.1	0.4	116	9.7	13.5	1.0
HTI	64.0	9.7	5.6	27.6	13.8	1708.8	179.3	1728.9	40.0	41.1	0.6	91	11.3	11.5	0.3
IDN	71.7	13.6	8.2	33.8	4.7	11459.1	151.2	11812.2	17.7	39.0	0.5	121	270.6	8.8	2.3
IND	69.7	12.2	6.5	30.2	5.4	6681.3	180.4	6754.3	25.7	37.8	0.5	109	1366.4	11.8	2.0
IRK	70.6	11.3	7.3	12.9	12.8	10800.8	119.9	10881.2	19.4	29.5	0.6	113	39.3	22.0	5.3
IRL	82.3	18.7	12.7	43.8	4.9	68370.6	106.6	86781.4	7.2	32.8	0.1	150	4.9	12.0	8.1
IRN	76.7	14.8	10.3	34.7	11.4	12446.7	550.9	14535.9	11.3	40.8	0.5	129	82.9	13.4	8.8
ISL	83.0	19.1	12.8	20.1	2.8	54682.4	129.0	55874.0	5.6	26.8	0.1	145	0.3	24.3	10.8
ISR	83.0	16.2	13.0	21.8	3.9	40186.8	108.2	40161.9	10.9	39.0	0.1	152	8.5	23.0	7.7
ITA	83.5	16.1	10.4	18.0	9.9	42776.4	110.6	42412.7	11.8	35.9	0.1	139	60.6	18.8	5.6
JOR	74.5	11.4	10.5	18.4	14.7	9858.5	125.6	9906.1	14.6	33.7	0.5	116	10.1	15.5	2.4
JPN	84.6	15.2	12.9	24.3	2.3	42931.7	105.5	41429.3	8.1	32.9	0.1	111	126.9	19.8	9.1
KAZ	73.6	15.6	11.9	27.0	4.6	22857.5	182.8	26351.4	7.1	27.5	0.2	136	18.6	9.2	17.6
KEN	66.7	11.3	6.6	17.4	2.6	4243.6	180.5	4329.9	26.2	40.8	0.5	97	52.6	13.1	0.4
KIR	71.5	13.0	11.1	32.9	6.3	4864.4	155.7	5253.1	9.5	27.7	0.4	119	6.4	16.8	1.6

LAO	67.9	11.0	5.3	29.0	0.6	7413.3	135.9	7826.4	24.7	36.4	0.5	116	7.2	14.0	2.7
LBR	64.1	9.6	4.8	22.8	2.8	1258.4	223.1	1427.8	31.8	35.3	0.7	97	4.9	19.7	0.3
LET	75.3	16.2	13.0	22.1	6.5	30282.4	116.9	30830.1	9.2	35.6	0.2	128	1.9	18.2	3.7
LKA	77.0	14.1	10.6	27.4	4.2	12707.4	155.5	13078.1	13.8	39.8	0.4	116	21.3	9.4	1.1
LSO	54.3	11.3	6.5	32.1	23.4	3150.6	155.9	2767.7	27.4	44.9	0.6	93	2.1	38.3	1.3
LTU	75.9	16.6	13.1	16.7	6.4	35798.7	118.4	36974.9	10.0	37.3	0.1	138	2.8	17.0	4.8
LUX	82.3	14.3	12.3	17.4	5.4	72711.7	115.1	114481.5	9.6	34.9	0.1	128	0.6	17.1	15.9
MCN	75.8	13.6	9.8	34.1	17.8	15864.6	113.4	16506.1	11.8	34.2	0.1	120	2.1	15.0	3.5
MDA	71.9	11.5	11.7	26.3	5.5	13663.6	166.2	13033.9	10.3	25.7	0.2	94	4.0	13.7	1.3
MEX	75.1	14.8	8.8	21.4	3.4	19159.8	141.5	19746.4	20.8	45.4	0.3	130	127.6	11.6	3.8
MGO	69.9	14.2	10.3	35.9	6.0	10838.7	195.8	12309.8	14.0	32.7	0.3	102	3.2	12.2	8.9
MLI	59.3	7.5	2.4	22.5	7.2	2268.8	108.7	2327.4	32.4	33.0	0.7	135	19.7	15.6	0.2
MLT	82.5	16.1	11.3	20.0	3.5	39554.5	113.5	43340.2	7.9	29.2	0.2	137	0.4	17.3	3.6
MNG	76.9	15.0	11.6	31.1	14.9	21399.2	116.3	21378.6	9.4	39.0	0.1	140	0.6	18.3	3.2
MOZ	60.9	10.0	3.5	43.9	3.2	1250.4	182.3	1280.5	30.7	54.0	0.5	95	30.4	22.9	0.3
MRT	64.9	8.6	4.7	40.9	9.5	5134.6	135.0	5197.0	31.8	32.6	0.6	125	4.5	12.9	0.6
MUC	75.0	15.1	9.5	20.0	6.7	25266.2	129.9	22989.3	13.6	36.8	0.3	126	1.3	15.3	3.8
MUI	64.3	11.2	4.7	12.3	5.7	1034.7	418.3	1059.7	28.6	44.7	0.6	111	18.6	13.9	0.1
NAM	63.7	12.6	7.0	12.7	20.3	9356.7	158.0	9637.2	33.6	59.1	0.4	106	2.5	25.9	1.7
NER	62.4	6.5	2.1	30.5	0.5	1200.9	109.3	1219.1	27.4	34.3	0.6	121	23.3	15.2	0.1
NIC	74.5	12.3	6.9	17.1	6.8	5283.7	162.7	5407.1	23.2	46.2	0.4	114	6.5	15.6	0.9
NOR	82.4	18.1	12.9	29.0	3.3	66494.3	120.3	63633.1	6.0	27.0	0.0	133	5.4	24.4	8.3
NPL	70.8	12.8	5.0	56.6	1.4	3456.7	188.7	3416.8	24.9	32.8	0.5	123	28.6	11.6	0.3
PAK	67.3	8.3	5.2	15.6	4.5	5005.0	182.3	4690.5	30.2	33.5	0.5	110	216.6	11.7	1.1
PAN	78.5	12.9	10.2	41.3	3.9	29557.7	122.1	31458.7	20.1	49.2	0.4	124	4.2	11.7	2.6
PBJ	82.3	18.5	12.4	21.2	3.2	57707.1	115.9	57141.4	6.9	28.5	0.0	126	17.1	24.4	9.5
PER	76.7	15.0	9.7	20.9	3.3	12252.3	129.8	12847.9	18.8	42.8	0.4	118	32.5	13.3	1.7
POL	78.7	16.3	12.5	19.6	3.5	31622.6	114.1	33086.4	7.6	29.7	0.1	140	37.9	17.9	9.1
PRT	82.1	16.5	9.3	18.9	6.3	33966.8	110.6	34798.0	11.8	33.8	0.1	139	10.2	16.9	5.0
PRY	74.3	12.7	8.5	22.4	4.8	12223.9	143.8	12684.6	22.8	46.2	0.4	115	7.0	11.8	1.1
ROM	76.1	14.3	11.1	22.9	4.0	29497.2	123.8	29909.0	11.4	36.0	0.3	140	19.4	17.3	3.8
RUA	69.0	11.2	4.4	26.1	1.0	2155.3	151.1	2226.2	28.4	43.7	0.4	97	12.6	16.2	0.1
RUS	72.6	15.0	12.2	23.1	4.6	26157.0	180.8	27043.9	10.0	37.5	0.2	137	145.9	18.3	11.7
SDN	65.3	7.9	3.8	19.3	16.5	3828.7	1344.2	3958.5	34.3	34.2	0.5	114	42.8	5.9	0.5

SEN	67.9	8.6	3.2	32.8	6.6	3309.4	109.3	3394.9	31.2	40.3	0.5	112	16.3	13.6	0.7
SLE	54.7	10.2	3.7	17.4	4.4	1667.8	234.2	1718.4	34.5	35.7	0.6	104	7.8	9.2	0.1
SLV	73.3	11.7	6.9	19.1	4.1	8359.1	111.2	8776.0	21.1	38.6	0.4	117	6.5	16.0	1.1
SRB	76.0	14.7	11.2	23.6	12.7	17191.7	144.0	18179.8	12.1	36.2	0.1	111	8.8	17.1	5.2
SUD	64.1	13.8	10.2	17.6	28.2	12129.2	158.9	12481.8	31.2	63.0	0.4	120	58.6	21.3	8.1
SUE	82.8	19.5	12.5	25.2	6.5	54507.8	110.5	53205.1	6.5	28.8	0.0	127	10.0	26.0	4.1
SUI	83.8	16.3	13.4	22.3	4.6	69393.5	99.5	68627.8	6.8	32.7	0.0	134	8.6	11.9	4.3
TAI	77.2	15.0	7.9	23.9	0.8	17781.2	113.3	18463.1	16.7	36.4	0.4	112	69.6	16.1	4.2
TAJ	71.1	11.7	10.7	27.2	11.0	3953.6	148.6	3379.7	12.4	34.0	0.3	92	9.3	17.8	0.6
TGO	61.0	12.7	4.9	28.0	2.0	1602.3	113.3	1595.9	31.7	43.1	0.6	108	8.1	18.2	0.4
TUN	76.7	15.1	7.2	19.3	16.0	10414.1	155.3	10755.6	18.9	32.8	0.3	149	11.7	20.6	2.7
TUR	77.7	16.6	8.1	25.1	13.5	27700.9	234.4	28167.4	16.5	41.9	0.3	149	83.4	15.7	5.2
TZA	65.5	8.1	6.1	34.0	2.0	2599.8	187.4	2660.4	24.9	40.5	0.6	109	58.0	8.5	0.2
UCR	72.1	15.1	11.4	12.6	8.9	13215.9	281.7	12810.3	6.5	26.1	0.2	118	44.0	19.9	5.1
UGA	63.4	11.4	6.2	26.5	1.8	2123.3	173.9	2181.2	26.7	42.8	0.5	90	44.3	7.9	0.1
UNG	76.9	15.2	12.0	28.6	3.4	31328.8	121.6	32622.9	7.3	30.6	0.2	127	9.7	19.4	5.1
URY	77.9	16.8	8.9	16.2	8.7	20063.6	202.9	21561.1	12.6	39.7	0.3	131	3.5	15.0	2.0
USA	78.9	16.3	13.4	21.0	3.7	63825.7	117.2	62682.8	12.1	41.4	0.2	148	329.1	14.1	16.6
VNM	75.4	12.7	8.3	26.8	2.0	7432.7	163.5	8041.2	16.5	35.7	0.3	127	96.5	6.5	2.2
ZAM	63.9	11.5	7.2	39.2	11.4	3325.5	212.3	3479.5	30.6	57.1	0.5	90	17.9	12.4	0.3
ZBE	61.5	11.0	8.5	9.3	5.0	2665.6	105.5	2835.9	22.5	44.3	0.5	84	14.6	25.6	0.8

Anexo 2

Componentes principales

	PC1	PC2	PC3
ALB	-0.66	-0.06	-0.59
ALG	-0.14	1.10	-0.48
ALM	-4.00	0.25	0.19
ARG	-0.98	-0.85	-0.89
ARM	-0.33	-1.27	-0.70
AUS	-4.63	-0.02	-0.13

AUT	-3.84	0.42	-0.02
BEL	-4.49	0.01	0.03
BEN	3.53	0.66	0.70
BFA	3.54	0.50	0.55
BGD	2.32	1.94	-0.36
BGR	-0.92	-0.38	0.64
BIH	-1.04	-1.62	-0.37
BOL	1.23	-0.45	0.81
BRA	0.73	-1.69	-0.14
CAN	-4.07	-0.09	-0.03
CHD	5.39	1.05	0.74
CHI	-2.15	-0.38	0.43
CHL	-0.91	-0.32	0.12
CHN	-0.30	4.38	-0.83
CHQ	-3.41	0.73	0.34
CMF	3.60	0.52	0.51
CMR	3.33	0.35	0.69
COG	3.36	-0.54	0.35
COL	0.69	-0.74	0.02
COR	-3.47	1.03	-0.11
CRI	-0.05	-1.48	0.14
CRO	-1.96	-0.22	0.33
DIN	-4.26	-0.28	0.49
DOM	1.01	0.46	0.18
ECU	0.73	0.14	0.61
EGI	1.11	0.03	-1.79
ESP	-2.58	-1.05	-0.29
EST	-3.30	0.36	0.11
ESU	3.10	-3.67	0.30
EVN	-3.32	0.22	0.14
EVQ	-2.04	0.07	0.51
FIL	1.23	0.75	0.53
FIN	-4.24	-0.28	0.24
FRA	-3.34	-0.34	0.09

GAB	1.64	-1.16	-0.62
GBR	-3.30	-0.19	0.14
GEO	-0.28	-0.61	-0.27
GHA	2.58	0.74	-0.80
GMB	3.48	-0.42	0.00
GRC	-2.48	-2.10	-0.45
GTM	2.27	-0.13	0.42
HND	2.25	-0.24	0.43
HTI	4.20	-0.49	0.04
IDN	1.27	1.85	-0.24
IND	2.33	3.47	-1.03
IRK	1.10	-1.51	0.27
IRL	-4.60	2.12	-0.31
IRN	0.10	0.35	-3.16
ISL	-4.75	-0.08	0.14
ISR	-3.30	-0.34	0.30
ITA	-2.65	-0.73	-0.17
JOR	0.79	-1.11	-0.17
JPN	-2.91	0.54	0.62
KAZ	-2.37	1.33	-1.11
KEN	2.80	0.02	0.53
KIR	0.33	0.76	0.16
LAO	2.14	1.20	0.67
LBR	3.62	0.01	0.67
LET	-1.79	-0.35	0.31
LKA	0.51	0.83	0.01
LSO	3.07	-3.49	1.34
LTU	-2.38	-0.62	0.04
LUX	-4.91	-0.07	-0.04
MCN	-0.47	-0.43	-0.35
MDA	0.26	0.58	0.25
MEX	0.19	0.46	-0.02
MGO	0.19	1.10	-0.17
MLI	3.70	0.24	0.36



MLT	-2.62	0.28	0.14
MNG	-1.42	-0.68	-0.26
MOZ	3.94	0.55	1.54
MRT	3.13	1.31	0.07
MUC	-0.60	-0.25	0.10
MUI	3.54	-0.89	-1.29
NAM	2.57	-3.92	0.57
NER	3.67	1.48	0.95
NIC	1.78	-0.80	0.30
NOR	-4.71	0.40	0.48
NPL	2.19	3.28	0.13
PAK	2.99	0.54	-0.15
PAN	0.21	1.43	0.48
PBJ	-4.38	-0.18	0.55
PER	0.54	0.12	0.42
POL	-2.92	0.22	-0.02
PRT	-2.22	-0.18	-0.04
PRY	1.41	0.05	0.33
ROM	-1.31	0.23	0.18
RUA	2.91	0.42	1.10
RUS	-1.90	0.24	-0.38
SDN	4.65	-0.79	-8.74
SEN	3.34	0.78	0.67
SLE	4.23	0.21	-0.16
SLV	1.39	-0.10	0.67
SRB	-0.82	-0.90	-0.08
SUD	1.56	-4.07	-0.49
SUE	-4.06	-0.49	0.63
SUI	-3.96	0.62	-0.07
TAI	0.09	0.73	0.76
TAJ	1.14	-0.54	0.49
TGO	3.22	0.28	1.35
TUN	-0.27	-1.44	-0.72
TUR	-0.94	-0.56	-1.32

TZA	3.45	1.74	0.18
UCR	-0.99	-1.08	-0.92
UGA	3.43	0.97	0.55
UNG	-2.00	0.58	0.45
URY	-0.58	-0.81	-0.56
USA	-3.93	0.95	-0.57
VNM	0.80	1.64	-0.41
ZAM	3.74	-0.11	0.33
ZBE	2.57	-2.04	1.90

### Anexo 3

*Tabla de datos transformada con las nuevas variables construidas*

	pc1	pc2	pc3
ALB	-1764.04	-2590.87	554.65
ALG	133.46	-605.84	355.47
ALM	-31682.78	-188.41	-13156.61
ARG	857.71	19232.28	-2554.09
ARM	-1874.52	-13629.36	-125.55
AUS	-30858.25	2806.83	-4324.28
AUT	-12598.10	1128.23	-51351.76
BEL	-33982.04	-2675.61	-24.15
BEN	-96.27	-360.20	-2.17
BFA	508.73	605.32	162.72
BGD	1773.31	-2681.87	-1368.42
BGR	-13533.99	2342.71	-5255.72
BIH	1800.39	4859.64	-14699.87
BOL	-4216.82	-3230.83	2685.77
BRA	699.60	49.05	-1387.53
CAN	-5950.17	-9453.07	2210.77
CHD	-36.77	-179.72	91.99
CHI	-22475.80	51.88	-9644.75
CHL	990.95	21103.28	-2788.52

CHN	-2431.31	-16449.91	400.55
CHQ	-24734.47	2253.52	-3517.83
CMF	-1175.64	236.93	-4683.44
CMR	-2382.56	-351.07	-52.19
COG	-215.33	-349.11	-5.46
COL	3327.01	4086.60	2207.09
COR	16117.72	-23224.99	-11990.30
CRI	-11170.45	1885.03	-4152.59
CRO	3412.66	9154.84	-28080.99
DIN	-28402.91	-21896.94	18348.68
DOM	945.59	49.68	-1680.48
ECU	-1317.78	-2117.56	461.44
EGI	191.30	-656.77	400.55
ESP	-23744.32	-60.20	-9991.35
EST	1560.50	32345.79	-4151.35
ESU	-1260.42	-8149.54	202.15
EVN	-24247.51	2194.71	-3352.09
EVQ	-7385.16	723.12	-29415.76
FIL	-6160.05	-582.78	-49.43
FIN	111.36	-5462.23	187.33
FRA	10869.44	12132.71	7764.59
GAB	5194.64	-8142.93	-4191.13
GBR	-27134.03	4655.63	-10344.35
GEO	1810.11	4683.77	-14731.53
GHA	-2542.77	-2003.16	1660.08
GMB	-15.50	28.78	-275.23
GRC	-3737.47	-5787.47	1330.17
GTM	93.10	-486.84	274.23
HND	-3276.16	29.70	-1423.08
HTI	9.46	1561.68	-250.52
IDN	-1693.87	-11579.52	145.29
IND	-4748.48	416.62	-736.98
IRK	-2431.75	325.62	-9916.71
IRL	-50789.20	-4939.51	-90.19

IRN	-901.08	-1556.41	31.10
ISL	12665.65	15241.57	9116.60
ISR	15051.45	-21855.85	-11283.79
ITA	-24875.81	4301.82	-9621.38
JOR	1212.44	3212.81	-9774.42
JPN	-20685.21	-16052.07	13431.60
KAZ	2070.64	-15.29	-2156.35
KEN	-509.64	-731.84	117.98
KIR	118.25	-325.54	205.00
LAO	-4473.36	25.30	-1931.82
LBR	-27.38	1218.31	-213.25
LET	-4147.12	-30056.08	78.11
LKA	-8228.77	726.13	-1146.68
LSO	-553.70	183.90	-2885.56
LTU	-23824.09	-1997.54	-44.76
LUX	-13539.63	-11528.50	90.88
MCN	3695.98	4579.25	2519.57
MDA	5040.01	-7076.49	-3639.98
MEX	-11403.28	1941.68	-4307.81
MGO	1487.07	3464.53	-11912.62
MLI	-1080.78	-899.37	713.33
MLT	2927.31	25.04	-3672.78
MNG	-2681.12	-4030.50	888.00
MOZ	-58.16	-199.93	102.38
MRT	-3066.63	0.63	-1292.96
MUC	1026.61	21683.77	-2498.59
MUI	-316.24	-1113.25	-8.33
NAM	-6074.00	530.48	-851.80
NER	-261.25	162.09	-1148.20
NIC	-3507.83	-452.45	-66.58
NOR	1179.51	-7247.82	275.14
NPL	814.12	908.11	314.09
PAK	1781.87	-2675.30	-1363.84
PAN	-17903.75	3015.86	-6624.08

PBJ	6826.37	18887.48	-56524.35
PER	-6190.94	-4588.72	3848.64
POL	1804.37	73.27	-2964.45
PRT	-4077.18	-6701.42	1583.09
PRY	249.66	-629.77	372.52
ROM	-17237.27	-4.94	-7313.25
RUA	34.46	1981.53	-298.39
RUS	-3711.71	-26198.31	236.75
SDN	-2964.28	113.19	-367.05
SEN	-749.76	196.62	-3081.01
SLE	-1114.06	-309.39	-86.56
SLV	-184.27	-967.07	20.27
SRB	4017.00	5117.10	2743.14
SUD	4471.29	-6872.15	-3508.67
SUE	-31403.69	5461.81	-12265.60
SUI	8187.60	22715.01	-67894.85
TAI	-8964.30	-6642.00	5576.30
TAJ	-126.44	50.47	-434.92
TGO	-205.36	-203.14	-29.56
TUN	175.09	-578.51	340.95
TUR	-16268.79	11.16	-6915.59
TZA	39.28	2392.41	-355.75
UCR	-1808.53	-12922.98	-161.92
UGA	-1479.91	108.60	-208.30
UNG	-7410.22	761.46	-28748.71
URY	-13627.91	-1306.52	-83.32
USA	484.25	-7109.65	250.57
VNM	1763.96	2345.95	1043.58
ZAM	1140.18	-1933.92	-966.60
ZBE	-1605.59	268.56	-613.41