

▼ A2-Componentes Principales

Jorge Eduardo de Leon Reyna - A00829759

```
1 # Librerias
2 install.packages("factoextra")
3 install.packages("ggplot2")

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

▼ Parte 1

A partir de los datos sobre indicadores económicos y sociales de 96 países hacer un análisis de Componentes principales a partir de la matriz de varianzas-covarianzas y otro a partir de la matriz de correlaciones, comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

```
1 # Importando dataset
2 data = read.csv("/content/paises_mundo.csv")
3 head(data)
```

A data.frame: 6 × 11

	CrecPobl	MortInf	PorcMujeres	PNB95	ProdElec	LinTelf	ConsAgua	PropBosq	P
	<dbl>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	
1	1.0	30	41	2199	3903	12	94	53	
2	3.0	124	46	4422	955	6	57	19	
3	4.3	21	13	133540	91019	96	497	1	
4	2.5	34	24	44609	19883	42	180	2	
5	1.3	22	31	278431	65962	160	1043	22	
6	1.4	6	43	337909	167155	510	933	19	

▼ 1. Calculando matriz de covarianza y de correlacion

```
1 # Matriz de covarianza
2 covarianza_matriz = cov(data)
3 covarianza_matriz
```

A matrix: 11 x 11 c

	CrecPobl	MortInf	PorcMujeres	PNB95	ProdElec	
CrecPobl	1.538298e+00	2.195026e+01	-6.078026e+00	-8.933379e+04	-4.973964e+04	-1.3e
MortInf	2.195026e+01	1.032859e+03	-9.249342e+00	-2.269332e+06	-1.043435e+06	-4.3e
PorcMujeres	-6.078026e+00	-9.249342e+00	7.698322e+01	2.813114e+05	2.260248e+05	4.4e
PNB95	-8.933379e+04	-2.269332e+06	2.813114e+05	4.999786e+10	2.247791e+10	2.0e
ProdElec	-4.973964e+04	-1.043435e+06	2.260248e+05	2.247791e+10	1.821909e+10	7.5e
LinTelf	-1.369079e+02	-4.381366e+03	4.499750e+02	2.039550e+07	7.583050e+06	3.8e
ConsAgua	-4.827092e+01	-1.288211e+03	-1.568313e+03	1.097481e+07	1.399817e+07	1.1e
PropBosq	-3.887018e+00	-1.466316e+01	6.517895e+01	2.474311e+05	7.035979e+04	2.4e
PropDefor	3.361974e-01	1.276296e+01	2.680592e-01	-5.806203e+04	-3.180340e+04	-9.9e
ConsEner	-8.384169e+02	-4.442568e+04	2.855207e+02	1.415628e+08	6.801296e+07	3.4e
EmisCO2	-1.137877e+00	-9.485500e+01	-2.150132e+00	2.501673e+05	1.392779e+05	6.3e

```
1 # Matriz de correlacion
2 correlacion_matriz = cor(data)
3 correlacion_matriz
```

▼ 2. Calcule los valores y vectores propios de cada matriz

```

CrecPobl    1 000000000    0.55067948    -0.55852711  -0.32212154  -0.29711119  -0.56321228

1 # Valores y vectores propios de matriz de Covarianza
2 eigen_covarianza = eigen(covarianza_matriz)
3 valores_propios_covarianza = eigen_covarianza$values
4 vectores_propios_covarianza = eigen_covarianza$vectors
5

ProdElec   -0.29711119 -0.24053689    0.19085114  0.74476081  1.00000000  0.28664508

1 # Valores y vectores propios de matriz de correlacion
2 eigen_correlacion = eigen(correlacion_matriz)
3 valores_propios_correlacion = eigen_correlacion$values
4 vectores_propios_correlacion = eigen_correlacion$vectors

```

▼ 3. Calcule la proporción de varianza explicada por cada componente.

```

1 # Varrianza total explicada para matriz de covarianza
2
3 # Varianza total
4 varianza_total = sum(diag(valores_propios_covarianza))
5
6 # Varianza explicada
7 proporcion_varianza_explicada_covarianza = valores_propios_covarianza / varianza_total
8
9 print("Varianza explicada por componente")
10 print(proporcion_varianza_explicada_covarianza)

[1] "Varianza explicada por componente"
[1] 9.034543e-01 9.647298e-02 6.795804e-05 4.554567e-06 1.782429e-07
[6] 7.530917e-09 5.317738e-09 6.657763e-10 8.502887e-11 2.107843e-11
[11] 6.989035e-12

1 # Varrianza total explicada para matriz de correlacion
2
3 # Varianza total
4 varianza_total = sum(diag(valores_propios_correlacion))
5
6 # Varianza explicada
7 proporcion_varianza_explicada_correlacion = valores_propios_correlacion / varianza_total
8
9 print("Varianza explicada por componente")
10 print(proporcion_varianza_explicada_correlacion)

[1] "Varianza explicada por componente"
[1] 0.366352638 0.175453813 0.124582832 0.078592361 0.072194597 0.066290906
[7] 0.051936828 0.029709178 0.015278951 0.013302563 0.006305332

```

▼ 4. Acumulacion de resultados de varianza explicada

```

1 # Acumulacion de varianza con matriz de covarianza
2
3 acumulacion_prop_varianza_covarianza = cumsum(proporcion_varianza_explicada_covaria
4 print("Varianxa explicada acumulada")
5 print(acumulacion_prop_varianza_covarianza)

[1] "Varianxa explicada acumulada"
[1] 0.9034543 0.9999273 0.9999953 0.9999998 1.0000000 1.0000000 1.0000000
[8] 1.0000000 1.0000000 1.0000000 1.0000000

1 # Acumulacion de varianza con matriz de correlacion
2
3 acumulacion_prop_varianza_correlacion = cumsum(proporcion_varianza_explicada_correl
4 print("Varianxa explicada acumulada")
5 print(acumulacion_prop_varianza_correlacion)

[1] "Varianxa explicada acumulada"
[1] 0.3663526 0.5418065 0.6663893 0.7449816 0.8171762 0.8834671 0.9354040
[8] 0.9651132 0.9803921 0.9936947 1.0000000

```

5. Según los resultados anteriores, ¿qué componentes son los más importantes? ¿qué

- ▼ variables son las que más contribuyen a la primera y segunda componentes principales? ¿por qué lo dice? ¿influyen las unidades de las variables?

```

1 # Componentes principales matriz de covarianza
2
3 # Primer componente principal
4 primer_componente <- vectores_propios_covarianza[, 1]
5
6 # Segundo componente principal (corrección en el comentario)
7 segundo_componente <- vectores_propios_covarianza[, 2]

1 # Componentes principales matriz de correlacion
2
3 # Primer componente principal
4 primer_componente <- vectores_propios_correlacion[, 1]
5
6 # Segundo componente principal (corrección en el comentario)
7 segundo_componente <- vectores_propios_correlacion[, 2]

```

Solo se toman en cuenta los resultados de componentes principales de la matriz de covarianza al tener una variacion explicada mucho mayor.

1. Los dos primeros componentes son los más importantes, ya que juntos explican aproximadamente el 99.99% de la varianza total en los datos.
2. Las unidades de las variables si pueden influir en la magnitud de los valores en los vectores propios y por lo tanto en su contribución a los componentes principales.

▼ 6. Diferencias entre analisis de matriz de correlacion y covarianza

Tanto la varianza explicada como la varianza explicada acumulada es mucho mayor al hacer el correspondiente analisis en la matriz de covarianza por lo que se puede deducir que esta es la mejor opcion para hacer este tipo de analisis de componentes principales.

▼ Parte 2

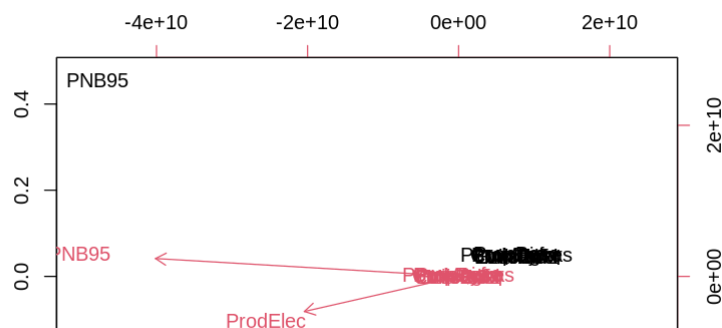
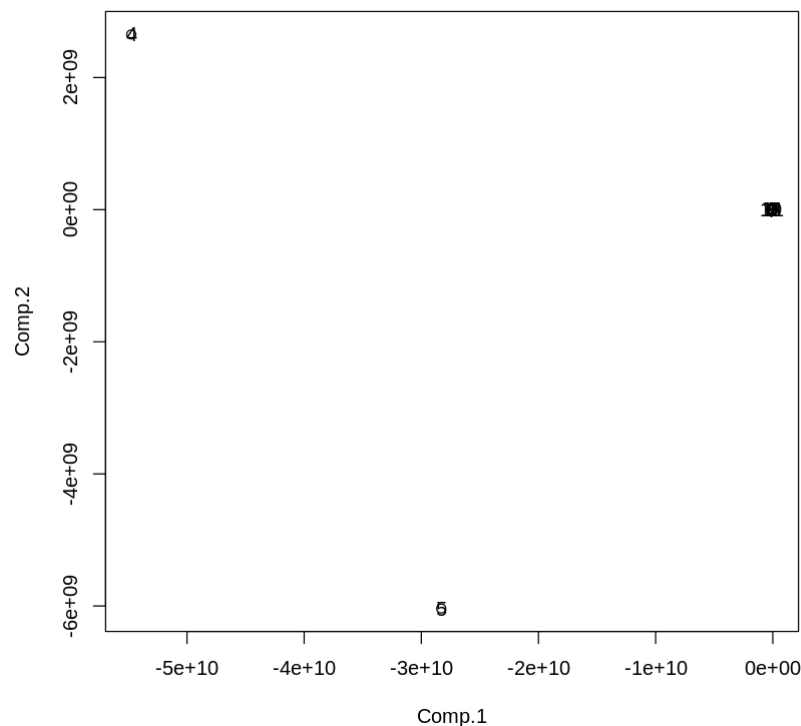
Obtenga las gráficas de respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes e interprete los resultados en término de agrupación de variables (puede ayudar "índice de riqueza", "índice de ruralidad")

```
1 library(stats)
2 library(factoextra)
3 library(ggplot2)

1 # Grafico para matriz de covarianza
2
3 datos = covarianza_matriz
4 cpS = princomp(datos,cor=FALSE)
5 cpaS = as.matrix(datos)%*%cpS$loadings
6 plot(cpaS[,1:2],type="p", main = "Título")
7 text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
8 biplot(cpS)
```

```
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
Warning message in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], len=
"zero-length arrow is of indeterminate angle and so skipped"
```

Título

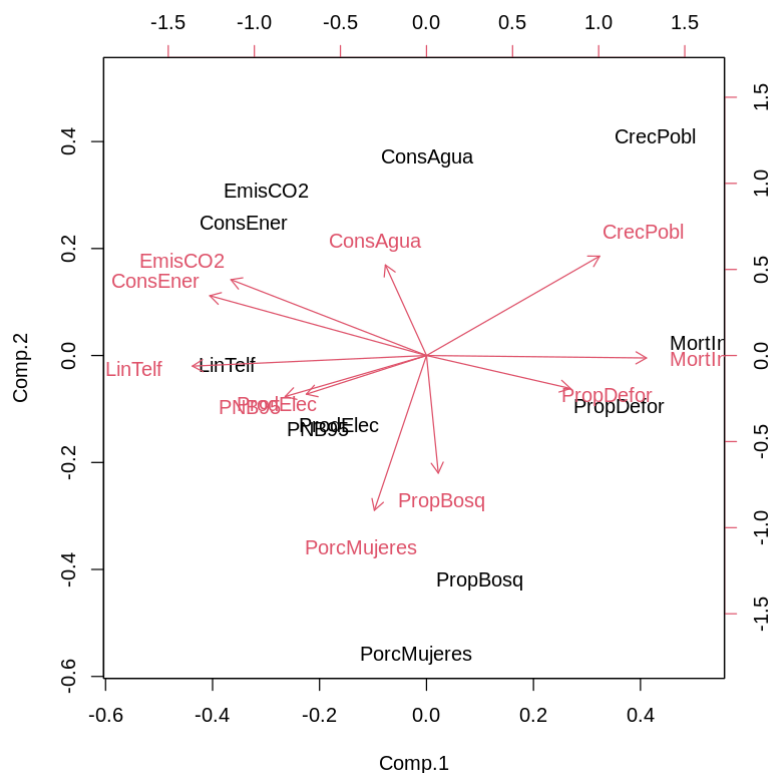
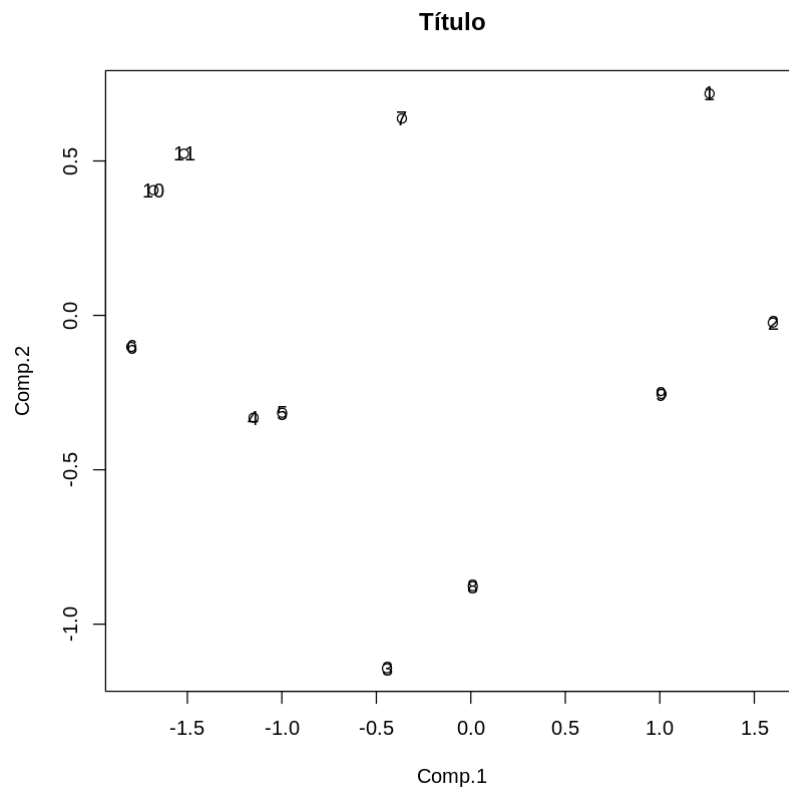


```
1 # Grafico para matriz de correlacion
2
3 datos = correlacion_matriz
4 cps = princomp(datos,cor=FALSE)
```

```

5 cpaS = as.matrix(datos)%*%cpS$loadings
6 plot(cpaS[,1:2],type="p", main = "Título")
7 text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
8 biplot(cps)

```



Análisis Al examinar detenidamente las gráficas generadas tanto para la matriz de covarianza como para la matriz de correlación, se destaca que es en la matriz de covarianza donde la identificación de componentes principales es más evidente. En esta matriz, con solo considerar los dos primeros componentes principales, podemos explicar una parte significativa de la varianza, como se demostró en el análisis anterior. En contraste, en la matriz de correlación, no es posible identificar componentes principales que expliquen la varianza de la misma manera que ocurre con la matriz de covarianza.

▼ Parte 3

Explore los siguientes gráficos relativos al problema y Componentes Principales y dé una interpretación de cada gráfico.

```
1 library(FactoMineR)
2
3
4 # Grafico matriz de covarianza
5 datos = covarianza_matriz
6 cp3 = PCA(datos)
7 grafico_pca = fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
8 ggsave("grafico_pca.png", plot = grafico_pca)
9
10 print(grafico_pca)
```

Warning message:

"ggrepel: 9 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
Saving 6.67 x 6.67 in image

```
1 # Grafico matriz de correlacion
2 datos = covarianza_matriz
3 cp3 = PCA(datos)
4 grafico_pca = fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
5 ggsave("grafico_pca.png", plot = grafico_pca)
6
7 print(grafico_pca)
```

Warning message:

"ggrepel: 9 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
Saving 6.67 x 6.67 in image

▼ Analisis final

1. Tras completar un análisis cuantitativo y gráfico exhaustivo, hemos identificado que las variables "PNB95" y "ProElec" se destacan como las más influyentes dentro de los componentes principales.
2. El análisis de componentes principales nos brinda la oportunidad de simplificar nuestros modelos al reducir la cantidad de variables involucradas, lo que, a su vez, simplifica la comprensión de los mismos.
3. Mediante un análisis gráfico cualitativo, hemos observado que la matriz de covarianza sobresale como la elección más efectiva para identificar los componentes principales en nuestros datos.
4. Es de suma importancia realizar un análisis tanto cualitativo como gráfico, ya que proporciona una comprensión global y múltiples perspectivas de los resultados obtenidos.