

Laboratorio 1: Análisis de componentes principales

KEVIN STEVEN GARCÍA^a, ALEJANDRO VARGAS^b

1. Introducción

En el presente informe veremos el uso y aplicación del análisis de componentes principales en una base de datos que contiene las importaciones hechas por los países suramericanos (Colombia, Brasil, Chile, Argentina, Ecuador y Perú), éstas provenientes de Estados Unidos, entre 1991 y 2010. Se analizará la cantidad de ejes o componentes principales a utilizar y se darán algunas interpretaciones de las trayectorias que se pueden formar entre los años consecutivos.

La base de datos sobre la cuál se va a trabajar es la siguiente:

Año	Colombia	Brasil	Chile	Argentina	Ecuador	Peru
1991	44.4	27.2	45.6	20.0	6.0	14.1
1992	75.5	11.8	58.9	22.6	17.8	14.4
1993	110.7	50.6	128.3	17.2	119.4	118.5
1994	80.3	70.6	102.2	15.2	154.9	146.1
1995	81.6	82.3	89.0	35.1	169.4	127.1
1996	76.4	97.4	185.0	51.0	75.5	129.0
1997	32.0	89.5	195.3	31.1	33.4	110.2
1998	55.5	63.1	66.3	24.4	9.7	66.7
1999	74.3	72.6	76.3	28.1	11.2	110.7
2000	84.5	76.2	80.1	29.5	11.8	110.2
2001	87.1	97.4	89.3	51.5	63.1	89.3
2002	89.3	89.5	72.4	40.3	66.3	70.2
2003	70.2	63.1	80.1	60.5	76.3	90.1
2004	90.1	66.3	70.5	39.1	20.0	64.5
2005	60.5	76.3	107.2	31.1	63.4	92.7
2006	140.3	20.0	63.4	50.2	101.2	120.8
2007	120.4	22.6	101.2	51.0	103.1	107.2
2008	130.2	17.2	103.1	42.5	66.7	70.8
2009	110.1	31.1	75.6	25.7	110.7	101.2
2010	120.2	24.4	68.9	60.3	110.2	110.8

^aCódigo: 1533173. E-mail: kevin.chica@correounivalle.edu.co

^bCódigo: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co

2. Procedimiento ACP

Para llevar a cabo el análisis de componentes principales de la base de datos anterior, estandarizaremos las variables que en nuestro caso son los países, ya que aunque todas están medidas en la misma escala, podría haber diferencias por el tipo de economía que maneja cada país. La matriz de correlaciones correspondiente a esta base de datos es:

	Colombia	Brasil	Chile	Argentina	Ecuador	Perú
Colombia	1	-0.524830476	-0.22431855	0.391517899	0.48620724	0.2790988
Brasil	-0.524830476	1	0.46610474	0.005930606	-0.04478948	0.3709440
Chile	-0.22431855	0.46610474	1	0.051451487	0.15920217	0.4937184
Argentina	0.391517899	0.005930606	0.05145149	1	0.16157351	0.1828657
Ecuador	0.4862072	-0.04478948	0.15920217	0.16157351	1	0.6547563
Perú	0.2790988	0.3709440	0.4937184	0.1828657	0.6547563	1

Para llevar a cabo el análisis de forma multivariada, debemos diagonalizar la matriz de correlaciones, es decir, obtener sus valores y sus vectores propios correspondientes.

Esta matriz de correlaciones tiene 6 valores propios positivos, que son:

$$\lambda_1 = 2.1934092$$

$$\lambda_2 = 1.9561781$$

$$\lambda_3 = 0.9038789$$

$$\lambda_4 = 0.5119470$$

$$\lambda_5 = 0.2854407$$

$$\lambda_6 = 0.1491461$$

Con los valores propios obtenidos, procedemos a hallar los vectores propios correspondientes

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
-0.3266572	0.5708984	0.0068014	0.1170875	0.5332445	0.5189071
-0.1588587	-0.6033277	0.2009878	-0.5413708	0.1852719	0.4929051
-0.3381710	-0.4635960	0.0514265	0.7890499	-0.0779079	0.1985066
-0.2898593	0.2066273	0.8863988	-0.0448699	-0.2455986	-0.1589085
-0.5459391	0.1756044	-0.3785943	-0.2216216	-0.6626128	0.1990175
-0.6096157	-0.1470297	-0.1669623	-0.1395665	0.4193810	-0.6192861

Ahora, para encontrar las componentes principales, hacemos el producto de la matriz de datos estandarizados con la matriz que contiene los vectores propios obtenidos de la matriz de correlaciones:

$$C = Z \cdot V$$

?????

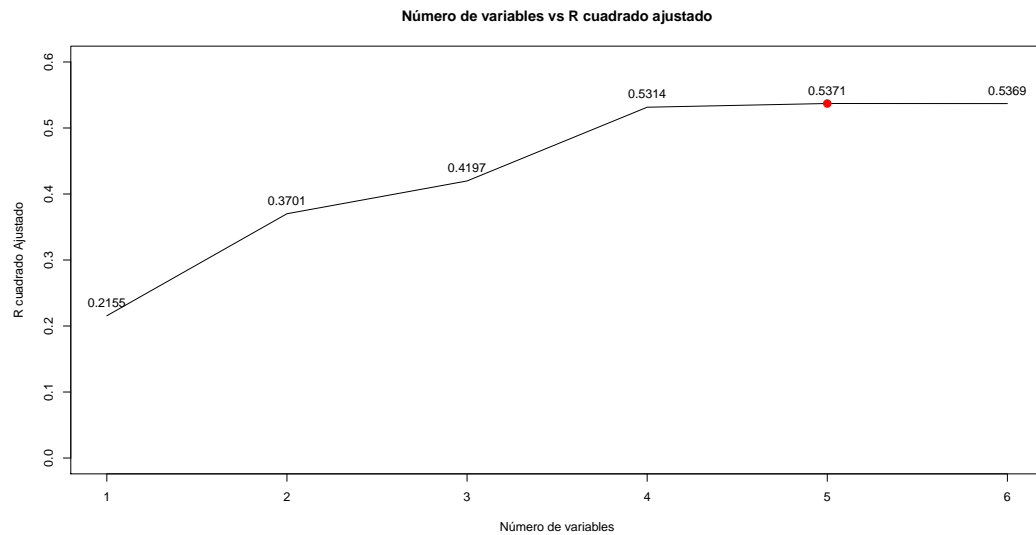


FIGURA 1: Gráfico de número de variables contra $R^2_{Ajustado}$