

Aplicación del Análisis de Componentes Principales a Variables Económicas y Vinculadas al COVID-19

Análisis Multivariado

Emanuelle Marsella, Maximiliano Saldaña

Junio 2021

- Aplicamos ACP a una base de datos sobre variables económicas y sanitarias en el contexto del COVID-19.
- Se tuvo como referencia el artículo **Economic Policy Responses to a Pandemic: Developing the COVID-19 Economic Stimulus Index** (Ceyhun Elgin, Gokce Basbug, Abdullah Yalaman; 2020). Crean un Índice de Estímulo Económico (*CESI_INDEX*) que buscamos replicar.
- Los autores utilizan el Índice como variable explicada en un modelo lineal, nosotros la pretendemos utilizar en un segundo ACP.
- Datos al 31 de marzo de 2020, fuente: FMI principalmente.

- Se cuenta con datos para 166 observaciones (países) y 14 variables.
- Las variables empleadas para replicar el *CESI_INDEX*:
 - *fiscal*: paquete de políticas fiscales adoptadas, representado como porcentaje del PBI
 - *ratecut*: representa el porcentaje que fue recortado de la tasa de interés en relación a la de febrero de 2020
 - *macrofin*: tamaño del paquete de medidas macrofinancieras, expresado como porcentaje del PBI
 - *bopgdp*: políticas de balanza de pago, expresadas como porcentaje del PBI
 - *othermonetary*: variable indicadora que indica si en el país se tomaron otras medidas monetarias
 - *otherbop*: variable indicadora que indica si en el país se tomaron otras medidas de políticas de balanza de pago.

Variables empleadas en el ACP posterior:

- *medage*: edad mediana en el país
- *infectionrate*: tasa de infección, representa la proporción de la población total infectada de COVID-19
- *hospitalbed*: camas de hospital por cada 1000 personas
- *healthexp*: gastos en salud, expresados como porcentaje del PBI
- *gdppercap*: PBI per cápita del país en dólares a niveles de 2010
- *CESI_INDEX*: índice de estímulo económico (construido por los autores utilizando análisis de componentes principales)

Variables no consideradas en el análisis final

- *totalcases*: cantidad de casos totales de COVID-19
- *stringency*: índice de rigurosidad de respuesta gubernamental (desarrollado por Hale y Webster en 2020)
- *denspob*: densidad poblacional del país, medida como cantidad de personas por kilómetro cuadrado de área terrestre.

	Variable	Datos faltantes
1	totalcases	15
2	medage	20
3	gdppercap	23
4	healthexp	20
5	hospitalbed	19
6	stringency	93
7	infectionrate	22

Table 1: Tabla de cantidad de observaciones faltables para las variables.

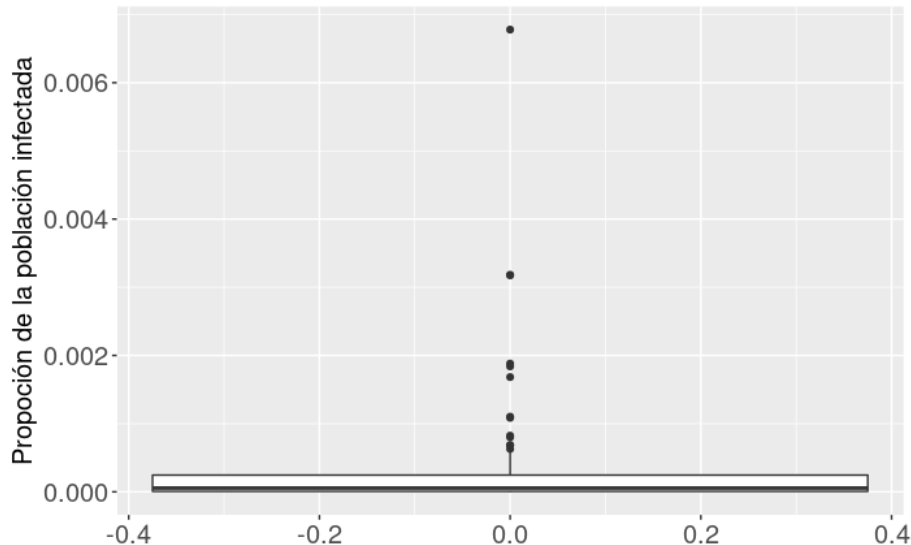
Correlaciones entre las variables

- Se destacan por ser altas las correlaciones entre:
 - *gdppercap* y *fiscal*: 0.6
 - *medage* y *gdppercap*: 0.6
 - *medage* y *hospitalbed*: 0.71.

Principales resultados del análisis descriptivo

- Argelia toma el valor más alto de *bopgdp* (6%, el país que le sigue es Croacia con 2.94%, para la mayoría de los países es 0%) y el más bajo de *fiscal* (-7.2%)
- 21 observaciones con valores altos de *ratecut*, destacándose particularmente Estados Unidos, que tuvo una reducción del 100% de su tasa de interés. La mediana de *ratecut* es 0%, 95 países toman este valor.
- En *macrofin* 102 países toman el valor 0 de esta variable.
- En *gdppercap* la observación que toma el mayor valor es Luxemburgo, con un PBI per cápita de 110742 USD.
- En *healthexp*, se destaca una la observación correspondiente a Estados Unidos con un gasto en salud del 17.07% del PBI, casi el doble que el país que le sigue.

Boxplot de *infectionrate*



Principales resultados del análisis descriptivo

- *infectionrate* cuenta una gran acumulación de tasas de infección en torno a valores próximos a 0, y un 75% de las tasas son menores a 0,025% (que representa un caso por cada 4000 habitantes aproximadamente).
- Se destaca especialmente San Marino, con una tasa de infección del 0,68% que representa un caso por cada 150 habitantes aproximadamente.

Construcción del Índice de Estímulo Económico

- Se usa *FactomineR* para recrear el Índice que los autores del artículo de referencia emplearon.
- Se pudo comprobar que el Índice de Estímulo Económico de los países que se obtuvo realizando el ACP y definiéndolo como los valores de la primera componente principal es equivalente al obtenido por los autores. Esto nos lleva a pensar que el procedimiento de imputación de valores faltantes fue el mismo, la imputación de la media de las variables.
- Al realizar el ACP observamos que la primera componente principal logra explicar un 27.13% de la variabilidad total de la nube de puntos, la segunda un 19.76% y la tercera un 18.04%.
- El porcentaje explicado explicado por la primera componente es bajo pero acorde al uso que le dan los autores.

Coordenadas de las variables en el primer eje

Variable	Correlación
fiscal	0.74
ratecut	0.50
macrofin	0.71
bopgdp	-0.21
othermonetary	0.50
otherbop	-0.21

Table 2: Correlación de cada variable con el índice de estímulo económico.

ACP considerando el Índice de Estímulo Económico

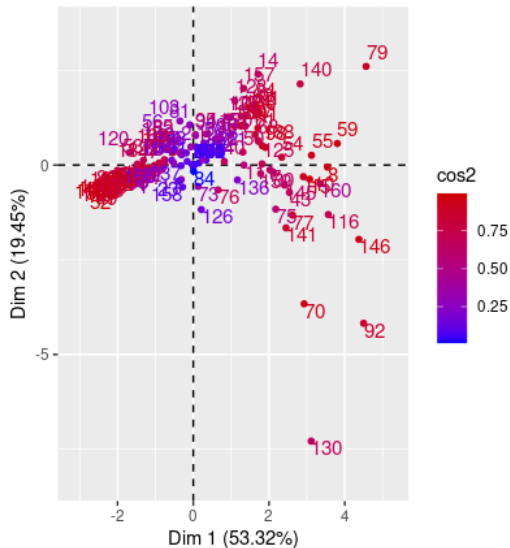
- Como primer paso, creamos a partir de *infectionrate* la variable *noinfectionrate*, para indicar la proporción de habitantes del país no infectados en lugar de los infectados.
- El conjunto de variables empleado en este análisis esta conformado por *medage*, *gdppercap*, *healthexp*, *hospitalbed*, *noinfectionrate*.
- La variable *CESI_INDEX*, el índice de estímulo económico obtenido a partir del ACP es utilizado como variable suplementaria en el análisis, y tampoco se incluyen las variables empleadas para construirlo al ser la intención utilizar el índice como su representante.

ACP considerando Índice de Estímulo Económico

	Valor propio	% de varianza	% acumulado de varianza
comp 1	2.67	53.32	53.32
comp 2	0.97	19.45	72.77
comp 3	0.67	13.49	86.26
comp 4	0.48	9.56	95.81
comp 5	0.21	4.19	100.00

Table 3: Valores propios asociados a las componentes principales y porcentajes de inercia de cada uno.

Perspectiva de la nube de individuos



Perspectiva de la nube de individuos

- Es deseable que ninguno de los individuos sea especialmente influyente a la hora de generar los componentes principales. A primera vista la figura anterior indicaría que esta situación no se da.
- El primer eje tiende a dar mayores valores a países de mayor desarrollo económico (que se corresponden con valores mayores en el conjunto de variables que se usaron para caracterizarlos).
- Mientras tanto el segundo eje tiende a dar valores mayores a países con menor tasa de infección.

Perspectiva de la nube de individuos

	País	cos2_plano	norma
130	San Marino	0.67	9.69
92	Luxembourg	0.85	6.64
79	Japan	0.90	5.55
160	United States	0.47	4.94
146	Switzerland	0.95	4.92
70	Iceland	0.94	4.83
116	Norway	0.73	4.46
140	South Korea	0.77	4.05
59	Germany	0.97	3.91
14	Belarus	0.64	3.69

Table 4: Observaciones con las 10 mayores normas y $\cos^2(\theta)$.

San Marino es una observación influyente, al considerarla como complementaria se gana un 4% de inercia total explicada. Estados Unidos tiene un bajo $\cos^2(\theta)$ pero no es influyente.

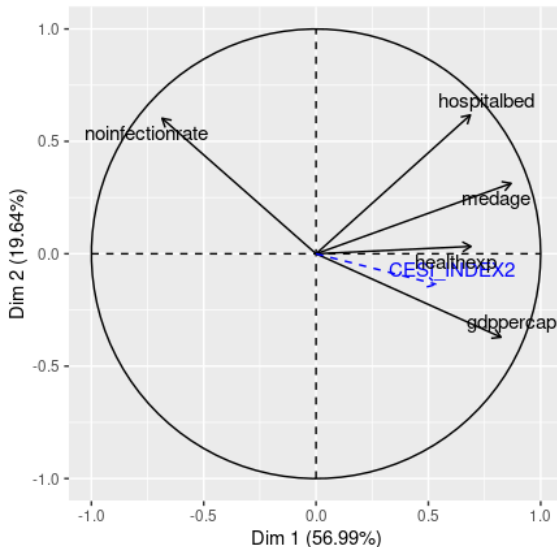
Perspectiva de la nube de individuos

	País	cos2_plano	norma
84	Kuwait	0.01	1.64
10	Bahamas	0.06	0.65
21	Botswana	0.06	0.00
26	Burundi	0.06	0.00
27	Cabo Verde	0.06	0.00
36	Congo, R	0.06	0.00
38	Cote Ivory	0.06	0.00
45	Ecuador	0.06	0.00
51	Eswatini	0.06	0.00
64	Guinea Bissau	0.06	0.00

Table 5: Observaciones con los 10 menores $\cos^2(\theta)$ y sus normas.

Se destaca que prácticamente todos son países en vías de desarrollo. No obstante, todos estos países se encuentran muy cercanos al baricentro, por lo que su influencia es limitada.

Perspectiva de la Nube de Variables



Perspectiva de la Nube de Variables

	Dim.1	Dim.2
medage	0.87	0.31
gdppercap	0.82	-0.37
healthexp	0.69	0.03
hospitalbed	0.69	0.62
noinfectionrate	-0.69	0.60

Table 6: Coordenadas de las variables originales en el primer plano principal

Perspectiva de la Nube de Variables

- Las correlaciones del índice de estímulo económico con los dos ejes principales, son 0.53 y -0.14, con el primer y segundo eje respectivamente. Esto significa que dicha variable se ve más relacionada con el eje que a su vez representa mejor a las variables *medage*, *gdppercap* y *healthexp*.
- Aún así debe ser tenido en cuenta que la correlación no resulta muy alta con el primer eje, lo que compromete la interpretación de la variable en el subespacio considerado.

Perspectiva de la Nube de Variables

Variable	Correlación
medage	0.49
gdppercap	0.58
healthexp	0.31
hospitalbed	0.22
noinfectionrate	-0.22

Table 7: Correlación del índice de estímulo económico construido con las variables originales de la base.

Conclusiones

- Se logró recrear con éxito el Índice de Estimulo Económico.
- Inicialmente en el ACP posterior se logró una inercia explicada del 72.77% con el plano principal.
- San Marino resultó ser un país influyente en el análisis y al excluirlo de las observaciones activas la inercia total explicada por los primeros dos ejes aumentó a 76.63%.
- Si bien todas las variables necesitan de ambas componentes consideradas para ser explicadas, la primera explica más el conjunto de las variables que expresan características económicas, del sistema de salud y demográficas; la segunda la tasa de no infección y la cantidad de camas de hospital cada 1000 habitantes.
- La proyección del índice de estímulo se comporta como el primer grupo de variables, aunque su correlación con la primer componente no resulta muy elevada en comparación con las variables que sí formaron parte del ACP.

Conclusiones

- De acuerdo a las direcciones y las correlaciones de las variables proyectadas en el primer plano principal, donde la tasa de no infección es opuesta a las variables socioeconómicas, los países con un mayor nivel de desarrollo socioeconómico fueron aquellos que presentaron altas tasas de infección.
- Esto representa la situación al 31 de marzo de 2020, cuando muchos de los países desarrollados se enfrentaban a una ola de infecciones.
- En consecuencia, vemos la dirección similar del índice de estímulo económico con las variables sociodemográfica: países más desarrollados, que fueron de los primeros en infectarse, tendieron a ser los que se vieron forzados a tomar medidas de estímulo económico para intentar enfrentarse a los estragos de la pandemia.

- Jorge Blanco. (2006). *Introducción al Análisis Multivariado*. Montevideo: IESTA.
- C. Elgin, G. Basbug y A. Salaman. (2020). *Economic Policy Responses to a Pandemic: Developing the COVID-19 Economic Stimulus Index*. CEPR Press.
- Marco Scavino. (2020). *Análisis Multivariado I - Análisis Factorial de una nube de puntos en un espacio vectorial euclidiano*. Montevideo: IESTA.