

# LOS ÍNDICES DE COLOMBIA

Santiago Serrano<sup>1</sup>

<sup>1,2</sup>Escuela de Ingeniería, Universidad de los Andes, [s.serrano11@uniandes.edu.co](mailto:s.serrano11@uniandes.edu.co)

05 de Julio de 2018

## Abstract

El proposito principal de este trabajo es describir los procesos para partir una poblacion N-dimensional en partes de k tamaño en forma de una muestra. El proceso, que se denomina 'k-means' aparece para dar particiones que son razonablemente eficientes en el sentido de las variables dentro de las categorias. Eso es, que si  $p$  es la probabilidad de densidad para la poblacion  $S$ ,  $S = \{s_1, s_2, \dots, s_n\}$ . La parte de  $E_n$  y de  $u_i$  siendo  $i = 1, 2, 3, \dots, k$  es el promedio condicional de  $p$  sobre  $S$ . Diremos de ahora en adelante en el documento 'tiende a ser bajo' para referirnos en principio a las consideraciones intuitivas y corroboradas del analisis matematico y practicas computacionales.

## Introducción

The main purpose of this paper is to describe a process for partitioning an N-dimensional population into k sets on the basis of a sample. The process, which is called 'k-means', appears to give partitions which are reasonably efficient in the sense of within-class variance. That is, if  $p$  is the probability mass function for the population,  $S = \{S_1, S_2, \dots, S_n\}$  is a partition of  $E_n$ . We say 'tends to be low' primarily because of intuitive considerations, corroborated to some extent by mathematical analysis and practical computational experience.

Comencemos viendo que hay en la sección 1 en la página 2.

# 1 Exploración Univariada

En esta sección exploro cada índice. En esta sección exploro cada índice. En esta sección exploro cada índice. En esta sección exploro cada índice. En esta sección exploro cada índice. En esta sección exploro cada índice. En esta sección exploro cada índice.

Para conocer el comportamiento de las variables se ha preparado donde se describe la distribución de las modalidades de cada variable. Los números representan la situación de algún país en ese indicador, donde el mayor valor numérico es la mejor situación.

Table 1: Medidas estadísticas

Statistic	Mean	St. Dev.	Max	Min
IDH	0.802	0.042	0.879	0.691
Poblaci..n.Cabecera	1,196,730.000	1,982,287.000	10,070,801	13,090
Poblaci..n.Resto	360,590.300	331,887.600	1,428,858	21,926
Poblaci..n.Total	1,557,320.000	2,202,522.000	10,985,285	43,446

Como apreciamos en las regiones de Colombia en la mejor situación son los menos, salvo en el caso del *índice de libertas mundial*<sup>1</sup>

---

<sup>1</sup>Nótese que esto se puede deber a la **menor** cantidad de categorías.

Para resaltar lo anterior, tenemos la Figura 1 en la página 3.

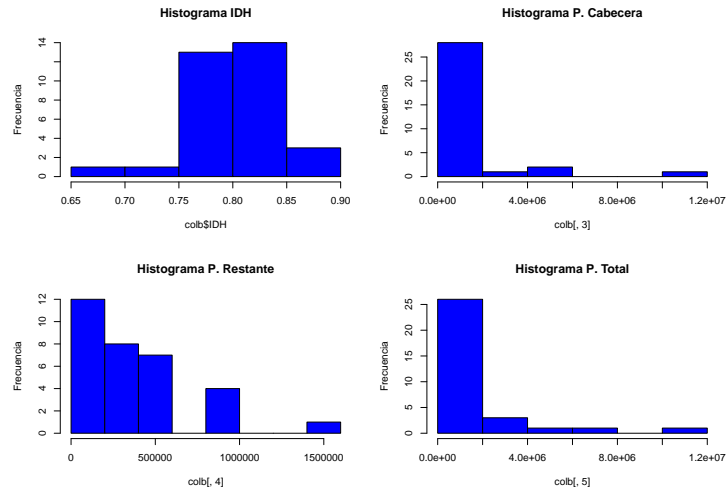


Figure 1: Histograma de Indicadores por categoría

Además de la distribución de los variable, es importante saber el valor central. Como los valores son de naturaleza ordinal debemos pedir la **mediana** y otras medidas de posición (como los *cuartiles*, los que no pediremos pues son pocos valores). La mediana de cada variable la mostramos en la página.

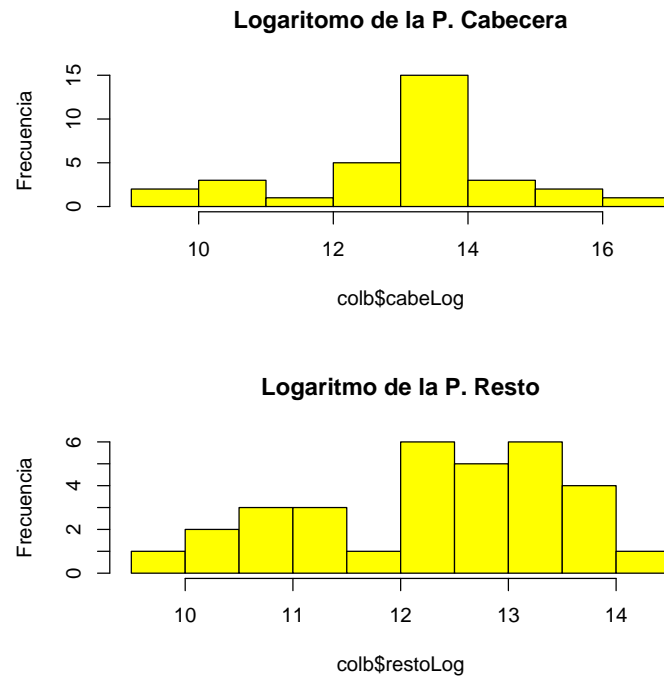


Figure 2: Histogramas de los logaritmos (transformación)

## 2 Exploración Bivariada

En este trabajo estamos interesados en el impacto de los otros índices en el nivel de Desarrollo Veamos las relaciones bivariadas que tiene esta variable con todas las demás:

Table 2: Correlacion de IDH con las demas variables

cabeLog	restoLog
0.487	0.177

Veamos la correlación entre las variables independientes:

Table 3: Correlacion entre las variables independientes

	cabeLog	restoLog
cabeLog	1	
restoLog	0.84	1

Lo visto en la Tabla 4 se refuerza claramente en la Figura 2.

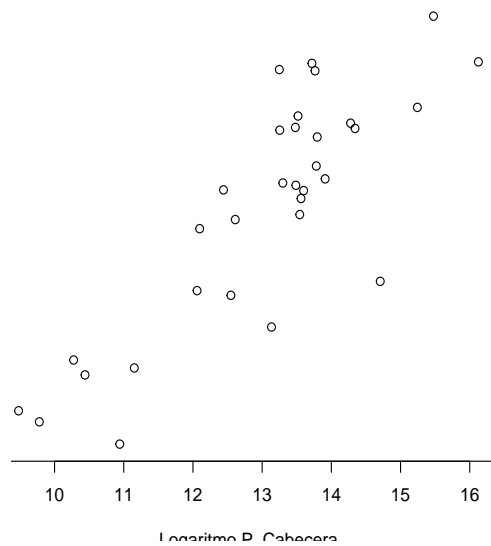


Figure 2: correlación entre predictores

### 3 Modelos de Regresión

El análisis de regresión se usa en estadística para calcular las relaciones entre variables. Generalmente en una variable dependiente y varias independientes.

Los resultados se muestran en la Tabla 4 de la página 6.

Table 4: Modelos de Regresion

	<i>Dependent variable:</i>	
	IDH	
	(1)	(2)
cabeLog	0.013*** (0.004)	0.031*** (0.007)
restoLog		-0.030*** (0.010)
Constant	0.634*** (0.055)	0.766*** (0.065)
Observations	32	32
R <sup>2</sup>	0.238	0.425
Adjusted R <sup>2</sup>	0.212	0.385
Residual Std. Error	0.037 (df = 30)	0.033 (df = 29)
F Statistic	9.347*** (df = 1; 30)	10.706*** (df = 2; 29)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## 4 Exploración Espacial

Para realizar la Exploración Espacial se utilizará la técnica de k-means propuesta por McQueen [1] para identificar conglomerados de las regiones de Colombia de la siguiente forma:

Se dividió el mapa de Colombia en un conjunto de  $n$  observaciones en  $k$  grupos. De esta forma se quiere dividir el mapa en tres categorías dependiendo del valor medio del desarrollo. Se clasificaron en tres categorías: alto, medio y bajo.

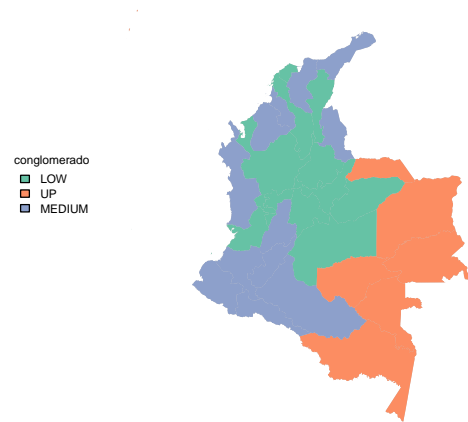


Figure 3: Mapa del IDH

## Bibliografía