

LOS INDICES DE COLOMBIA

Laura Valeria Vanegas García¹

¹Facultad de Ingeniería, Universidad de los Andes, lv.vanegas10@uniandes.edu.co

29 de Junio de 2018

Abstract

Este es mi proyecto final para el curso de verano IIN4347 de la Universidad de los Andes para 2018-19. Este es mi primer trabajo en exploración y modelamiento de índices usando LATEX. Este trabajo lo he hecho bajo la filosofía de trabajo replicable. Este es mi proyecto final para el curso de verano IIN4347 de la Universidad de los Andes para 2018-19. Este es mi primer trabajo en exploración y modelamiento de índices usando LATEX. Este trabajo lo he hecho bajo la filosofía de trabajo replicable.

Introducción

En este documento presento mi investigación sobre diversos índices sociales en Colombia. Los índices fueron extraídos de wikipedia y se realizó preprocesamiento en Python con ayuda de la librería Pandas. En este documento presento mi investigación sobre diversos índices sociales en Colombia. Los índices fueron extraídos de wikipedia y se realizó preprocesamiento en Python con ayuda de la librería Pandas.

En este documento presento mi investigación sobre diversos índices sociales en Colombia. Los índices fueron extraídos de wikipedia y se realizó preprocesamiento en Python con ayuda de la librería Pandas. En este documento presento mi investigación sobre diversos índices sociales en Colombia. Los índices fueron extraídos de wikipedia y se realizó preprocesamiento en Python con ayuda de la librería Pandas. En este documento presento mi investigación sobre diversos índices sociales en Colombia. Los índices fueron extraídos de wikipedia y se realizó preprocesamiento en Python con ayuda de la librería Pandas.

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

1 Exploración Univariada

Esta es la sección de exploración univariada. En esta sección exploro cada índice. Esta es la sección de exploración univariada. En esta sección exploro cada índice. Esta es la sección de exploración univariada. En esta sección exploro cada índice.

Además de la distribución de la 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 Tabla 1 en la página 2.

Table 1: Medidas estadísticas

Statistic	N	Median	Min	Max
IDH	32	0.804	0.691	0.879
Población.Cabecera	32	717,197	13,090	10,070,801
Población.Resto	32	268,111.5	21,926	1,428,858

A continuación, se muestra el histograma en la Figura 1 en la página 2.de las variables IDH, Población cabecera y Población Restante:

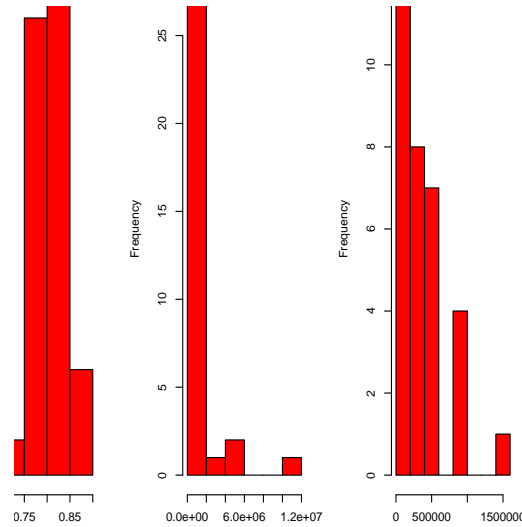


Figure 1: Histogramas

Dado el sesgo de las poblaciones, podríamos transformarla para que se acerque a la normalidad. En la Figura 2 en la página 3, se muestran los histogramas con las poblaciones normalizadas con una transformación de logaritmo.

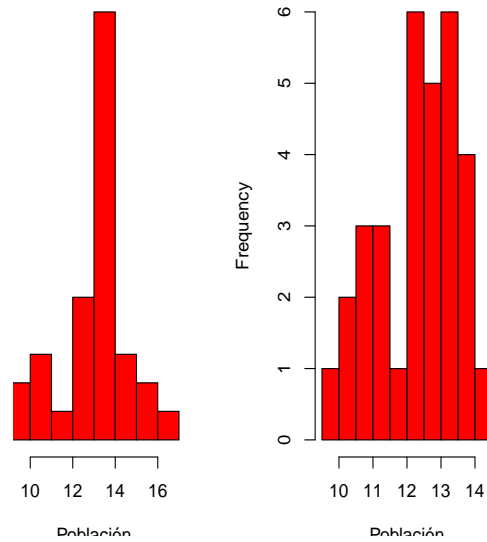


Figure 2: Histogramas Población Normalizados

2 Exploración Bivariada

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

Table 2: Correlación de Democracia con las demás variables

cabeLog	restoLog
0.487	0.177

Veamos la correlación entre las variables independientes:

Table 3: Correlación entre variables independientes

	cabeLog	restoLog
cabeLog	1	
restoLog	0.8	1

Lo visto en la Tabla 3 se refuerza claramente en la Figura 3.

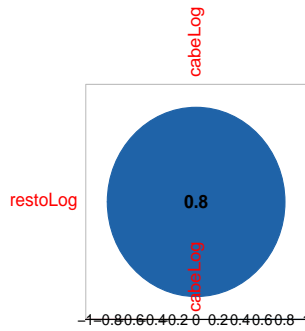


Figure 3: Correlación entre predictores

3 Modelos de Regresión

Finalmente, vemos los modelos propuestos. Primero sin poblacion resto, luego con esta. Los resultados se muestran en la Tabla 4 de la página 5.

Podemos sacar conclusiones al observar los resultados de las regresiones en

4 Exploración Espacial

Calculemos conglomerados de regiones, usando toda la información de las tres variables. Usaremos la tecnica de k-means propuesta por [1].

References

- [1] J. MACQUEEN. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. *MULTIVARIATE OBSERVATIONS*, page 17.

Table 4: Modelos de Regresión

	<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 ²	0.238	0.425
Adjusted R ²	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)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		