



Herramientas Computacionales para la Investigación Interdisciplinar Reproducible Proyecto Final

Nicolás Huertas - 201326730¹

¹Departamento de Ingeniería Industrial, Universidad de los Andes,
n.huertas10@uniandes.edu.col

30 de Junio de 2018

Abstract

Este proyecto busca explorar herramientas computacionales como Python, R, Zotero y LATEX para la generación de trabajo replicable. En este caso se explorarán variables de desarrollo humano para el caso Colombiano por departamento. De esta forma se identificará cuales son los departamentos que tienen mejores índices y peores así como ver la relación entre ellos. Para esto se hará un análisis univariado que permita identificar como se comporta cada índice luego

Introducción

Aquí les presento mi investigación sobre diversos índices sociales en el mundo. Los índices los conseguí de Wikipedia, espero que les gusten mucho. Aquí les presento mi investigación sobre diversos índices sociales en el mundo. Los índices los conseguí de Wikipedia, espero que les gusten mucho. Aquí les presento mi investigación sobre diversos índices sociales en el mundo. Los

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

1 Exploración Univariada

Para conocer el comportamiento de las variables se ha preparado la Tabla 1, 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

Table 1: Medidas estadísticas

Statistic	N	Mean	Median	Min	Max
IDH	32	0.802	0.804	0.691	0.879
Población.Cabecera	32	1,196,730.000	717,197	13,090	10,070,801
Población.Resto	32	360,590.300	268,111.5	21,926	1,428,858
Población.Total	32	1,557,320.000	1,028,429	43,446	10,985,285

valor numérico es la mejor situación. Como apreciamos en la Tabla 1, los países en la mejor situación son los menos, salvo en el caso del *índice de libertas mundial*¹

Para resaltar lo anterior, tenemos la Figura 1 en la página 2.

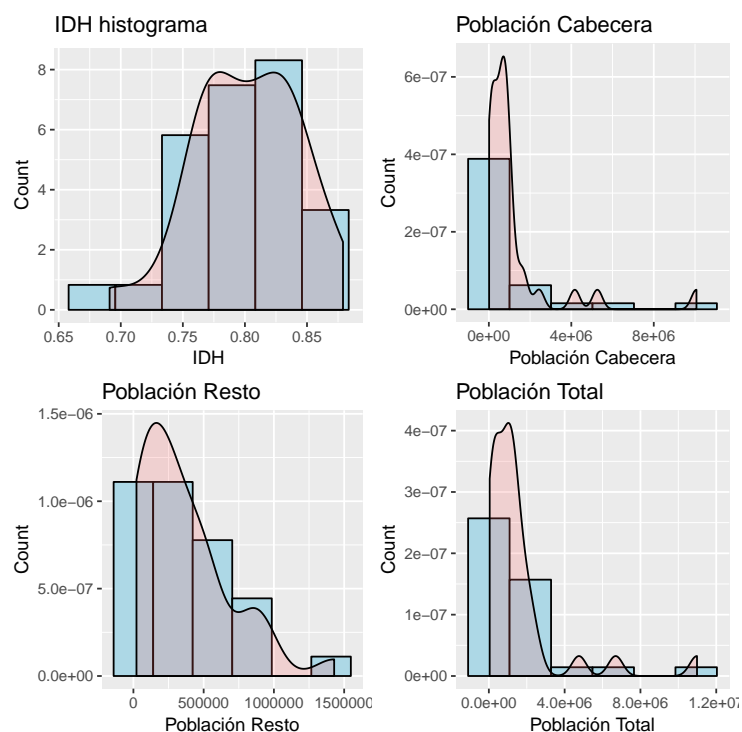


Figure 1: Distribución de Indicadores

dado el sesgo de las pobaciones, podriamos transformarla para que se acerque a la normalidad

¹Nótese que esto se puede deber a la **menor** cantidad de categorías.

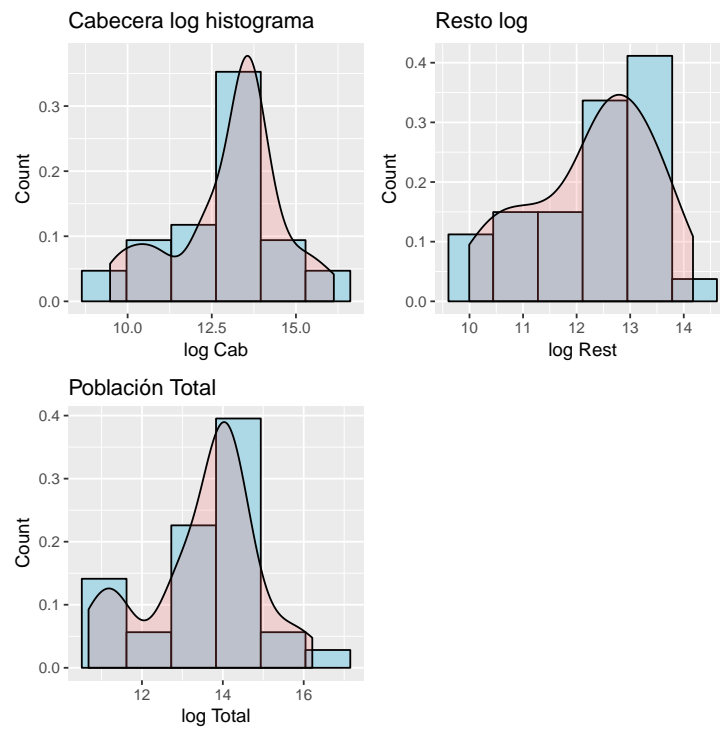


Figure 2: Distribución de Indicadores normalizada

2 Exploración Bivariada

En este trabajo estamos interesados en el impacto de la población en el nIDH. Veamos las relaciones bivariadas que tiene esta variable con todas las demás: Veamos la correlación entre las variables independientes:

Table 2: Correlación de IDH con las demás variables

cabeLog	restoLog	totalLog
0.487	0.177	0.424

Table 3: Correlación entre variables independientes

	cabeLog	restoLog	totalLog
cabeLog	1		
restoLog	0.84	1	
totalLog	0.99	0.9	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 la libertad mundial como independiente, y luego con está. Los resultados se muestran en la Tabla 4 de la página 5. Como se vió en la Tabla 4, cuando está presente el *índice de*

Table 4: Modelos de Regresión

	<i>Dependent variable:</i>	
	IDH	
	(1)	(2)
cabeLog	0.013*** (0.004)	0.066 (0.046)
restoLog		-0.016 (0.020)
totalLog		-0.051 (0.064)
Constant	0.634*** (0.055)	0.818*** (0.092)
Observations	32	32
R ²	0.238	0.437
Adjusted R ²	0.212	0.377
Residual Std. Error	0.037 (df = 30)	0.033 (df = 28)
F Statistic	9.347*** (df = 1; 30)	7.257*** (df = 3; 28)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

libertad mundial, el *índice de libertad de prensa* pierde significancia.

4 Exploración Espacial

Calculemos conglomerados de regiones, usando toda la información de las tres variables. usaremos la tecnica de k-means propuesta por Reynolds et al. (2006).

Como acabamos de ver en la Tabla 4 en la página 5, si quisieras sintetizar la multidimensionalidad de nuestros indicadores, podríamos usar tres de las cuatro variables que tenemos (un par de las originales tiene demasiada correlación).

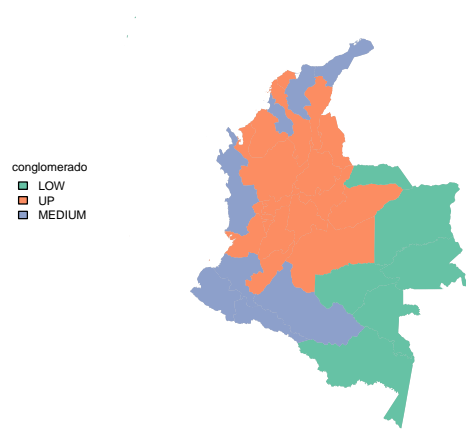


Figure 4: Regiones conglomerados segun sus poblaciones

Bibliografía

Reynolds, A. P., Richards, G., de la Iglesia, B., and Rayward-Smith, V. J. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, 5(4):475–504.