

PROYECTO FINAL

Andres Felipe Cardona Rodriguez¹

¹Facultad de Ingenieria Industrial, Universidad de los Andes, Curso: Herramientas Computacionales para la Investigacion Interdisciplinar Reproducible. IIND 4347,

¹Codigo: 201127229, Correo: af.cardona1552@uniandes.edu.co,

01 de Julio de 2018

Abstract

El presente articulo tiene como objetivo mostrar estadisticas de los departamentos que conforman el pais de Colombia. Las estadisticas estan enfocadas a analizar las variables de IDH (Indice de Desarrollo Humano), Poblacion Cabecera y Poblacion Resto.

Introduccion

El presente articulo surge como una investigacion desarrollada en el curso de Herramientas Computacionales para la Investigacion Interdisciplinar Reproducible dictado en la Universidad de los Andes por el profesor Jose Manuel Magallanes, PhD. Primero se realizara una Exploracion Univariada, luego una Exploracion Bivariada, para posteriormente hacer un analisis de Regresion y por ultimo hacer una Exploracion Espacial donde los datos son analizados a traves de clusters en el mapa presentado. Se comenzara con la seccion 1 en la pagina 2

```
'data.frame':      32 obs. of  6 variables:
 $ IDH              : num  0.879 0.867 0.865 0.849 0.842 0.839 0.837 0.835 0.834 0.832 ...
 $ Departamento    : chr  "Santander" "Casanare" "Valle del Cauca" "Antioquia" ...
 $ Poblacion.Cabecera: int  1587972 281548 4169553 5262172 742812 761658 10070801 2438533 56
 $ Poblacion.Resto   : int  502867 93701 586560 1428858 539251 206109 914484 107391 21926 68
 $ PoblacionTotal     : int  2090839 375249 4756113 6691030 1282063 967767 10985285 2545924 7
 $ DepartamentoNorm  : chr  "Santander" "Casanare" "Valle del Cauca" "Antioquia" ...
```

1 Exploracion Univariada

En esta seccion se encontrara todo lo referente a las estadisticas univariadas del IDH, Poblacion Cabecera y Poblacion Resto. Se evaluara la Media, la Mediana, la desviacion estandar, asi como los valores maximos y los minimos. Posteriormente se presentaran los histogramas para luego ser normalizados.

A continuacion se cuenta con las estadisticas de las 32 observaciones (de los Departamentos de Colombia) con respecto al IDH, a la Poblacion Cabecera y a la Poblacion Resto

Table 1: Medidas estadisticas

Statistic	N	Mean	St. Dev.	Median	Min	Max
IDH	32	0.802	0.042	0.804	0.691	0.879
Poblacion.Cabecera	32	1,196,730.000	1,982,287.000	717,197	13,090	10,070,801
Poblacion.Resto	32	360,590.300	331,887.600	268,111.5	21,926	1,428,858

Para resaltar lo anterior a continuacion se presentan las graficas..

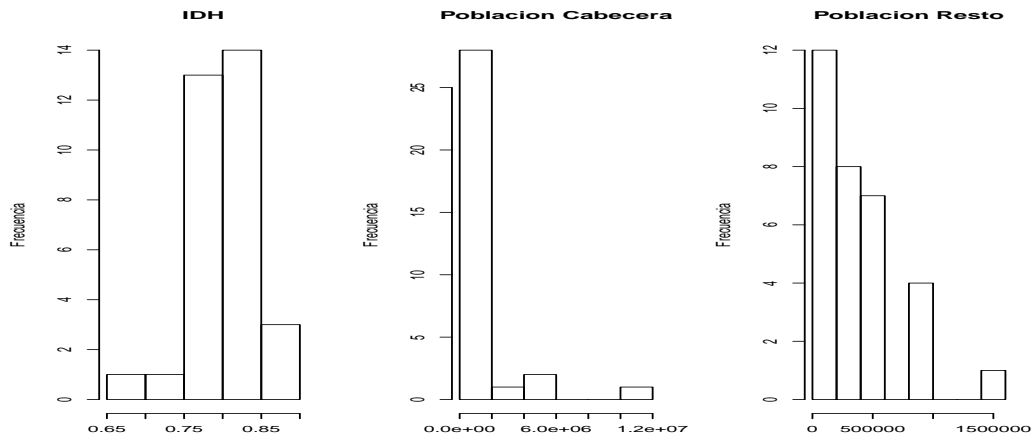


Figure 1: Histogramas IDH, Pobl. Cabecera y Pobl. Resto

Dado el sesgo de las poblaciones, podríamos transformarla para que se acerque a la realidad

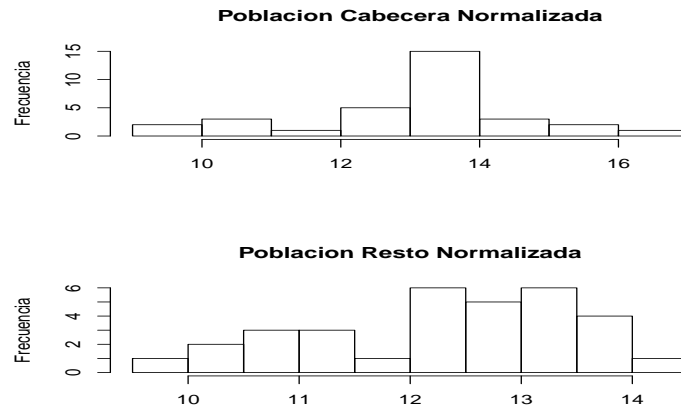


Figure 2: Histogramas normalizados

2 Exploration Bivariada

En esta seccion se vera la correlacion que tiene el IDH en relacion con las demas variables

En este trabajo estamos interesados en el impacto de la poblacion en el IDH, veamos IDH con cada uno:

Table 2: Correlacion del IDH con cada Poblacion

PobCabeceraLog	PobRestoLog
0.487	0.177

Y la correlacion entre las variables independientes:

Table 3: Correlacion entre variables independientes

	PobCabeceraLog	PobRestoLog
PobCabeceraLog	1	
PobRestoLog	0.84	1

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

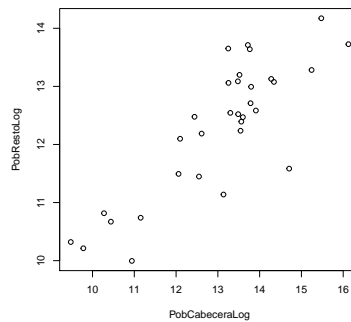


Figure 3: correlacion entre las variables independientes

3 Modelos de Regresion

Veamos los modelos propuestos, primero sin poblacion resto, luego con esta:

Table 4: Modelos de Regresion

	<i>Dependent variable:</i>	
	IDH	
	(1)	(2)
PobCabeceraLog	0.013*** (0.004)	0.031*** (0.007)
PobRestoLog		-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		

4 Exploracion Espacial

Calculemos conglomerados de regiones, usando toda la informaci??n de las tres variables. usaremos la tecnica de k-means propuesta por MacQueen.

[1]

[1] 1 3 2

	Group.1	IDH	PobCabeceraLog	PobRestoLog
1	1	0.7825714	10.58974	10.60684
2	2	0.7710833	13.28360	12.89513
3	3	0.8406154	14.12024	12.64415

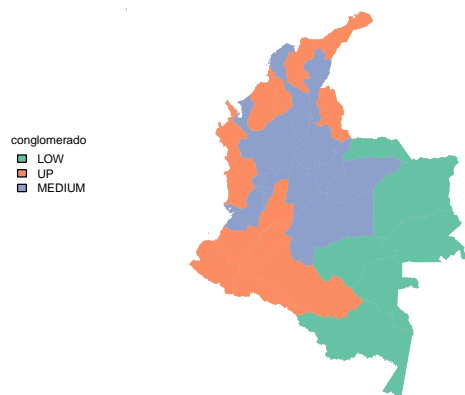


Figure 4: 3 Clusters Departamentos de Colombia

Bibliography

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