

LOS INDICES DE COLOMBIA

MACQUEEN¹

^{1,2}Escuela de Ingeniería, Universidad de los Andes,
delcurso, deallado@uniandes.edu.col

¹Instituto de altas investigaciones financieras, Banco del Parque,
delcurso@bp.com.col

30 de Junio de 2018

Abstract

El proposito principal de este trabajo es describir los procesos para partir una poblacion N-dimensional en partes de k tama<U+00F1>o en forma de una muestra. El proceso, que se denomina 'k-means' aparece para dar particiones que son rasonablemente eficientes en el sentido de las variables dentro de las categorias. Eso es, que si p es la probabilidad de densidad para la poblaci<U+00F3>n S, S=s1,s2,...,Sn. La parte de En y de ui siendo i=1,2,3,...,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 an<U+00E1>alisis matematico y practicas computacionales.

Introducción

The main purpose of this paper is to descrube a procces for partitioning an N-dimensional population into k sets on the basis of a sample. Th procvess, which is called 'k-means', appears to five partitions which are reasonably effcient in the sense of within-class variance. That is, if p is the probability mass functon for the population, S= S1,S2,...,Sn is a partition of En. We say 'tends to be low' primarily because of intutive considerations, corroborated to some extent by mthematical analysis and practical computational experience.

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

1 Exploraci??n Univariada

En esta sección exploro cada índice. En esta sección exploro cada índice. En
esta sección exploro cada índice. En esta secci??n exploro cada ??ndice. En
esta sección exploro cada índice. En esta sección exploro cada índice. En esta
sección exploro cada índice. En esta secció'n exploro cada índice. En esta sección
exploro cada índice.

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

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

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

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

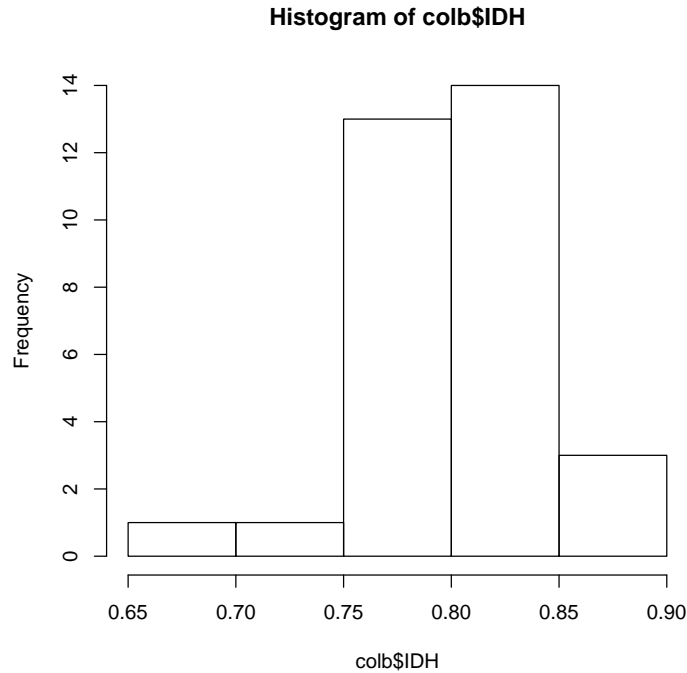


Figure 1: Distribución de Indicadores

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 Tabla ?? en la página ??.

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:

```
cabeLog restoLog [1,] 0.4873974 0.1773112
```

Veamos la correlación entre las variables independientes:

```
cabeLog restoLog cabeLog "1" "" restoLog "0.84" "1" Lo visto en la Tabla ??
```

se refuerza claramente en la Figura 2.

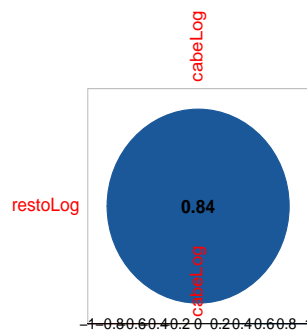


Figure 2: correlación entre predictores

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 ?? de la página ??.

Call: `lm(formula = IDH ~., data = colb[, c(1, 7)])`

Residuals: Min 1Q Median 3Q Max -0.113668 -0.018473 0.001249 0.016927 0.071401

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.634408 0.055163 11.501 1.6e-12 *** cabeLog 0.012846 0.004202 3.057 0.00466 ** — Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03737 on 30 degrees of freedom Multiple R-squared: 0.2376, Adjusted R-squared: 0.2121 F-statistic: 9.347 on 1 and 30 DF, p-value: 0.004664 Call: `lm(formula = IDH ~., data = colb[, c(1, 7:8)])`

Residuals: Min 1Q Median 3Q Max -0.09489 -0.02041 0.00433 0.01740 0.06372

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.765665 0.064813 11.813 1.32e-12 *** cabeLog 0.030664 0.006886 4.453 0.000116 *** restoLog -0.029571 0.009626 -3.072 0.004592 ** — Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03302 on 29 degrees of freedom Multiple R-squared: 0.4247, Adjusted R-squared: 0.3851 F-statistic: 10.71 on 2 and 29 DF, p-value: 0.0003296 Como se vió en la Tabla ??, cuando está presente el *índice de libertad mundial*, el *índice de libertad de prensa* pierde significancia.

4 Exploración Espacial

Como acabamos de ver en la Tabla ?? en la página ??, 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).

Así, propongo que calculemos conglomerados de países usando toda la información de tres de los indicadores. Como nuestras variables son ordinales utilizaremos un proceso de conglomeración donde las distancias serán calculadas usando la medida **gower** propuestas en [?], y para los enlazamientos usaremos la técnica de **medoides** según [?]. Los tres conglomerados se muestran en la Figura 3.

```
[1] "NAME_1"      "ID_0"        "ISO"         "NAME_0"      "ID_1"        "TYPE_1"
[7] "ENGTYPE_1"  "NL_NAME_1"  "VARNAME_1"  "IDH"         "cabeLog"     "restoLog"
```

Bibliografía

```

[1] "NAME_1"      "ID_0"      "ISO"      "NAME_0"    "ID_1"      "TYPE_1"
[7] "ENGTYPE_1"  "NL_NAME_1" "VARNAME_1" "IDH"      "cabeLog"   "restoLog"
[13] "cluster"

[1] 1 3 2

  Group.1      IDH  cabeLog restoLog
1      1 0.7825714 10.58974 10.60684
2      2 0.7710833 13.28360 12.89513
3      3 0.8406154 14.12024 12.64415

```

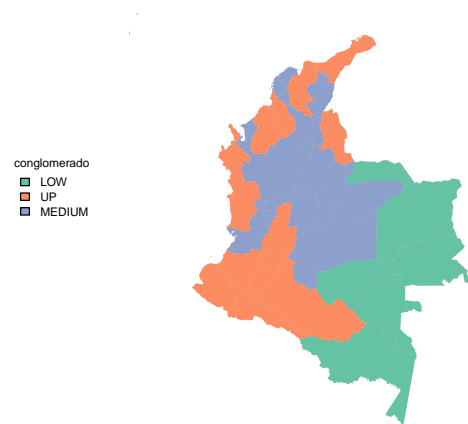


Figure 3: Países conglomerados segun sus indicadores sociopolíticos