LOS ÍNDICES DE COLOMBIA

Santiago Serrano[1]

[1,2]Escuela de Ingeniería,Universidad de los Andes, s.serrano11@uniandes.edu.col
[1]Instituto de altas investigaciones financieras, Banco del Parque,
delcurso@bp.com.col

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 tamano 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 poblacion 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 analisis 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 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 secci´n 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 algun país en ese indicador, donde el mayor valor numérico es la mejor situación.

Table 1: Medidas estad<U+00ED>sticas

Statistic	Mean	St. Dev.	Max	Min	Median
IDH	0.802	0.042	0.879	0.691	0.804
Poblaci..n.Cabecera	1,196,730.000	1,982,287.000	10,070,801	13,090	717,197
Poblaci..n.Resto	360,590.300	331,887.600	1,428,858	21,926	268,111.5
Poblaci..n.Total	1,557,320.000	2,202,522.000	10,985,285	43,446	1,028,429

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

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

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

■barplots, echo=FALSE,fig=TRUE■= par(mfrow=c(2,2))
title='IDH' paleta='red'
hist(colb$IDH, main = title, col = paleta, xlab =' IDH', ylab =' Frecuencia')$
title='Poblaci<U+00F3>n de cabecera' paleta='red'
hist(colb$poblacion.cabecera, main = title, col = paleta, xlab =' Poblaci <$
$U + 00F3 > n', ylab = "Frecuencia")$
title='Poblaci<U+00F3>n resto' paleta='red' hist(colb$Poblac, main =$
$title, col = paleta, xlab =' Poblaci < U + 00F3 > n', ylab = "Frecuencia")$
title='Poblaci<U+00F3>n total' paleta='red'
hist(colb$poblacion.total, main = title, col = paleta, xlab =' Poblaci <$
$U + 00F3 > n', ylab = "Frecuencia")$

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 enn la página.

■summary, fig=TRUE, echo=false■== dado el sesgo de las pobaciones, podriamos transformarla para que se acerque a la normalidad

par(mfrow=c(2,1))

colb$cabeLog = log(colb$poblacion.cabecera) colb$restoLog = log(colb$poblacion.resto)

title='Logaritmo de la poblaci<U+00F3>n de cabecera' paleta='red' demoTableRelPlot=hist(colb$cabeLog,
$title, col = paleta, xlab ='', ylab = "Frecuencia")$

title='Logaritmo de la poblaci<U+00F3>n restante' paleta='red' demoTableRelPlot=hist(colb$restoLog, m
$title, col = paleta, xlab ='', ylab = "Frecuencia")$

hist(colb$cabeLog)hist(colb$restoLog)

Histogramas de los logaritmos

2 Exploración Bivariada

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

■corrDem, results=tex, echo=FALSE■=

library(stargazer) explanans=names(colb)[c(7:8)] usando las logs corrDem=cor(colb$IDH, colb[, explanans]$, "na.or.complete")

stargazer(corrDem)

Veamos la correlación entre las variables independientes:

■corrTableX, results=tex, echo=FALSE■= y la correlaci<U+00F3>n entre las variables independientes:

corrTableX=round(cor(colb[,explanans], use = "na.or.complete"),2) corrTableX$_copy = corrTableXcorrTableX[upper.tri(corrTableX)] < −""$

ver: corrTableX

visualmente

stargazer(corrTableX,title="Correlaci<U+00F3>n entre las variables independientes")

Lo visto en la Tabla 1 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 **??**.

■regresiones, echo=FALSE■= LinRegA = lm(IDH ~ ., data = colb[,c(1,7)]) LinRegB = lm(IDH ~ ., data = colb[,c(1,7:8)]) stargazer(LinRegA,LinRegB,title="Modelos de Regresi<U+00F3>n",label="regresiones")

Como se vió en la Tabla **??**, cuando está presente el *indice 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 distancia 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.

■getMap, echo=FALSE,results=hide■= Exploraci<U+00F3>n Espacial —————————————————-

Calculemos conglomerados de regiones, usando toda la informaci<U+00F3>n de las tres variables. usaremos la tecnica de k-means propuesta por MacQueen.

library(rgdal) folder='COL$_m$$aps'file =' COL_adm1.shp'mapaFile = file.path(folder, file)mapCol <- rgdal :: readOGR(mapaFile, stringsAsFactors = F)$

lo tenemos: plot(mapCol)

veamos que variables hay:

head(mapCol@data)

con esto hagamos el merge: sub$_c$$olb = colb[, c(1 : 2, 7 : 8)]mapCol_idh = merge(mapCol, sub_colb, by.x =' NAME'_1, by.y =' Departamento', all.x = F)$

cuantas regiones me quedaron luego del merge? nrow(mapCol$_i$$dh)todas!!...

preparacion para clusterizar:

que tengo?: names(mapCol$_i$$dh)nombredelavariablesqueusar < U+00E9 >: dimensions = c("NAME_1","IDH","cabeLog","restoLog")$

creo un nuevo data frame con esas: dataCluster=mapCol$_i$$dh@data[, c(dimensions)]$

como la data es numerica la normalizo (menos la column 1): dataCluster[,-1]=scale(dataCluster[,-1])

APLICANDO TECNICA KMEANS

calculo 3 clusters

resultado=kmeans(dataCluster[,-1],3)

creo data frame con los clusters: clusters=as.data.frame(resultado$cluster)$

a<U+00F1>ado columna con nombre de regiones clusters$NAME_1 = dataCluster$NAME$_1$$names(clusters)c('cluster','NAME'_1)hagoelmergehaciaelmapa : mapCol_idh = merge(mapCol_idh, clusters, by =' NAME'_1, all.x = F)$

lo tengo? names(mapCol$_i$$dh)$

a pintar:

library(RColorBrewer) library(classInt)

variable a colorear varToPLot=mapCol$_i$$dhcluster$

decidir color: unique(varToPLot) aggregate(mapCol$_i$$dh@data[, c(10, 11, 12)], by = list(mapCol_idh@data$cluster),FUN=mean)

Bibliograf<U+00ED>a

■plotMap1, echo=FALSE, fig=TRUE■=
preparo colores numberOfClasses = length(unique(varToPLot))
colorForScale='Set2' paleta = brewer.pal(numberOfClasses, colorForScale)
grafico mapa basico plot(mapCol,col='grey',border=0)
grafico mapa cluster plot(mapCol$_i dh, col = paleta[varToPLot], border = F, add = T) legend('left', legend = c("LOW","UP","MEDIUM"), fill = paleta, cex = 0.6, bty = "n", title = "conglomerado")

Figure 3: Mapa del IDH