Informe segmentación

Jordi López Sintas

23/2/2017

# Introducción

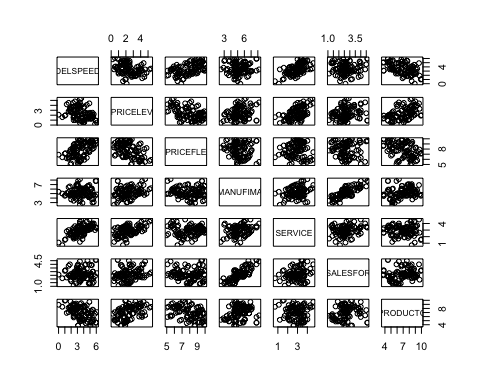
Descripción de las 6 primeras empresas y sus bases

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DELSPEED | PRICELEV | PRICEFLE | MANUFIMA | SERVICE | SALESFOR | PRODUCTQ |
| 4.1 | 0.6 | 6.9 | 4.7 | 2.4 | 2.3 | 5.2 |
| 1.8 | 3.0 | 6.3 | 6.6 | 2.5 | 4.0 | 8.4 |
| 2.7 | 1.0 | 7.1 | 5.9 | 1.8 | 2.3 | 7.8 |
| 4.6 | 2.4 | 9.5 | 6.6 | 3.5 | 4.5 | 7.6 |
| 2.4 | 1.6 | 8.8 | 4.8 | 2.0 | 2.8 | 5.8 |
| 2.8 | 1.4 | 8.1 | 3.8 | 2.1 | 1.4 | 6.6 |

Descripicón de las 6 primeras empreas

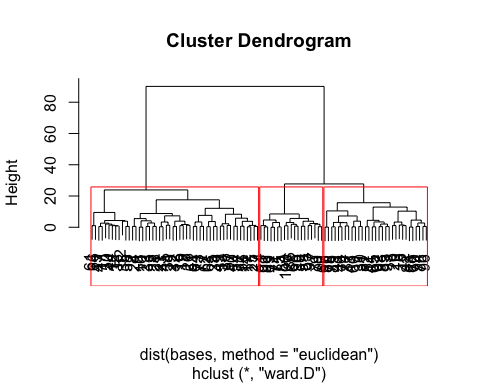
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TAMEMP | USAGELEV | SATISFLE | ESPCOMPR | ESTRCOMP | INDUSTRI | SITCOMP |
| Small | 32 | 4.2 | Total Value Analysis | Decentralized | Industry A Classification | New Task |
| Large | 43 | 4.3 | Specification Buying | Centralized | Other Industries | New Task |
| Large | 32 | 3.9 | Specification Buying | Centralized | Industry A Classification | New Task |
| Small | 46 | 5.8 | Total Value Analysis | Decentralized | Industry A Classification | New Task |
| Small | 32 | 4.3 | Total Value Analysis | Decentralized | Other Industries | New Task |
| Large | 39 | 4.4 | Specification Buying | Centralized | Other Industries | New Task |

# Descripción de Las bases



# Exploración de la heterogeneidad

## The "ward" method has been renamed to "ward.D"; note new "ward.D2"



Medias de la agrupación jerárquica

|  |  |  |
| --- | --- | --- |
| 1 | 2 | 3 |
| 4.32 | 2.57 | 4.69 |
| 1.58 | 3.15 | 1.57 |
| 8.66 | 6.89 | 9.29 |
| 4.59 | 5.57 | 5.47 |
| 2.91 | 2.84 | 3.12 |
| 2.18 | 2.82 | 3.05 |
| 5.08 | 8.04 | 7.24 |

# Partición de la muestra y clasificación

## Percepciones de las empresas en los segmentos

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| DELSPEED | 4.09 | 2.57 | 4.88 |
| PRICELEV | 1.62 | 3.21 | 1.51 |
| PRICEFLE | 8.63 | 6.80 | 9.37 |
| MANUFIMA | 4.42 | 5.60 | 5.81 |
| SERVICE | 2.83 | 2.87 | 3.18 |
| SALESFOR | 2.08 | 2.82 | 3.30 |
| PRODUCTQ | 5.27 | 8.13 | 7.00 |

# Descripción de las empresas que forman cada segmento

## Stratified by segmento  
## 1 2   
## n 33 48   
## TAMEMP = Small (%) 31 (93.9) 10 (20.8)   
## USAGELEV (mean (sd)) 46.33 (8.90) 42.73 (7.70)   
## SATISFLE (mean (sd)) 4.84 (0.74) 4.38 (0.76)   
## ESPCOMPR = Total Value Analysis (%) 31 (93.9) 10 (20.8)   
## ESTRCOMP = Decentralized (%) 31 (93.9) 0 ( 0.0)   
## INDUSTRI = Other Industries (%) 23 (69.7) 22 (45.8)   
## SITCOMP (%)   
## Modified Rebuy 8 (24.2) 22 (45.8)   
## New Task 8 (24.2) 24 (50.0)   
## Straight Rebuy 17 (51.5) 2 ( 4.2)   
## Stratified by segmento  
## 3 p test  
## n 19   
## TAMEMP = Small (%) 19 (100.0) <0.001   
## USAGELEV (mean (sd)) 54.21 (7.00) <0.001   
## SATISFLE (mean (sd)) 5.64 (0.57) <0.001   
## ESPCOMPR = Total Value Analysis (%) 19 (100.0) <0.001   
## ESTRCOMP = Decentralized (%) 19 (100.0) <0.001   
## INDUSTRI = Other Industries (%) 5 ( 26.3) 0.008   
## SITCOMP (%) <0.001   
## Modified Rebuy 2 ( 10.5)   
## New Task 2 ( 10.5)   
## Straight Rebuy 15 ( 78.9)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | p | test |
| n | 33 | 48 | 19 |  |  |
| TAMEMP = Small (%) | 31 (93.9) | 10 (20.8) | 19 (100.0) | <0.001 |  |
| USAGELEV (mean (sd)) | 46.33 (8.90) | 42.73 (7.70) | 54.21 (7.00) | <0.001 |  |
| SATISFLE (mean (sd)) | 4.84 (0.74) | 4.38 (0.76) | 5.64 (0.57) | <0.001 |  |
| ESPCOMPR = Total Value Analysis (%) | 31 (93.9) | 10 (20.8) | 19 (100.0) | <0.001 |  |
| ESTRCOMP = Decentralized (%) | 31 (93.9) | 0 ( 0.0) | 19 (100.0) | <0.001 |  |
| INDUSTRI = Other Industries (%) | 23 (69.7) | 22 (45.8) | 5 ( 26.3) | 0.008 |  |
| SITCOMP (%) |  |  |  | <0.001 |  |
| Modified Rebuy | 8 (24.2) | 22 (45.8) | 2 ( 10.5) |  |  |
| New Task | 8 (24.2) | 24 (50.0) | 2 ( 10.5) |  |  |
| Straight Rebuy | 17 (51.5) | 2 ( 4.2) | 15 ( 78.9) |  |  |

# Interpretación de los segmentos según percepciones y empresas

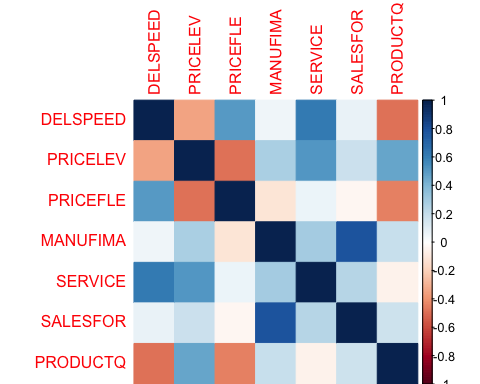
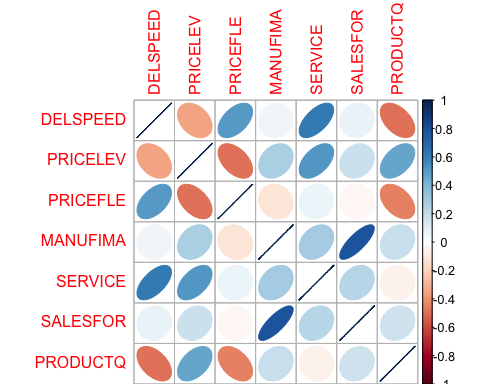
# Estrategia

# Anexo

Ahora comprobaríamos si el resultado obtenido con las bases de segmetnacion originales varía cuando las transformamos en sus componentes principales. Esta transformación nos permite comprobar que la correlación entre las bases de segmentación no suponga un problema para identificar los segmentos que queremos identificar.

Para ello primero inspeccionamos las correlaciones ente las bases.

## DELSPEED PRICELEV PRICEFLE MANUFIMA SERVICE  
## DELSPEED 1.00000000 -0.3492251 0.50929519 0.0504142 0.61190069  
## PRICELEV -0.34922515 1.0000000 -0.48721259 0.2721868 0.51298082  
## PRICEFLE 0.50929519 -0.4872126 1.00000000 -0.1161041 0.06661728  
## MANUFIMA 0.05041420 0.2721868 -0.11610408 1.0000000 0.29867737  
## SERVICE 0.61190069 0.5129808 0.06661728 0.2986774 1.00000000  
## SALESFOR 0.07711522 0.1862433 -0.03431610 0.7882245 0.24080818  
## PRODUCTQ -0.48263094 0.4697458 -0.44811201 0.1999811 -0.05516130  
## SALESFOR PRODUCTQ  
## DELSPEED 0.07711522 -0.4826309  
## PRICELEV 0.18624325 0.4697458  
## PRICEFLE -0.03431610 -0.4481120  
## MANUFIMA 0.78822454 0.1999811  
## SERVICE 0.24080818 -0.0551613  
## SALESFOR 1.00000000 0.1772939  
## PRODUCTQ 0.17729392 1.0000000

 Después realizamos algunas pruebas para ver si vale la pena estimar los componentes principales de las bases. El primero es el test de Barlett. Si podemos rechazar la hipótesis nula, entonces no es necesario estimar los componentes.

##   
## Bartlett test of homogeneity of variances  
##   
## data: bases  
## Bartlett's K-squared = 87.479, df = 6, p-value < 2.2e-16

El resultado del test de Barlett nos indica que podemos rechazar la hipótesis nula. Ahora comprobamos el test de adecuación de la muestra, KMO. Si su valor está cercano a 1, entonces es necesario estimar los componentes principales de las bases de segmentación.

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = bases)  
## Overall MSA = 0.45  
## MSA for each item =   
## DELSPEED PRICELEV PRICEFLE MANUFIMA SERVICE SALESFOR PRODUCTQ   
## 0.34 0.33 0.91 0.56 0.29 0.55 0.93

Como vemos el valor está pro debajo de 0,5, lo que nos indica que no es neceario continuar. No obstante continuaremos para mostrar que el resultado no cambia como sugieren los tests realizados.

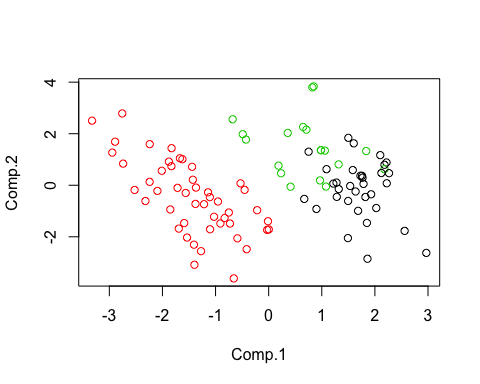
## Estimamos los componentes principales

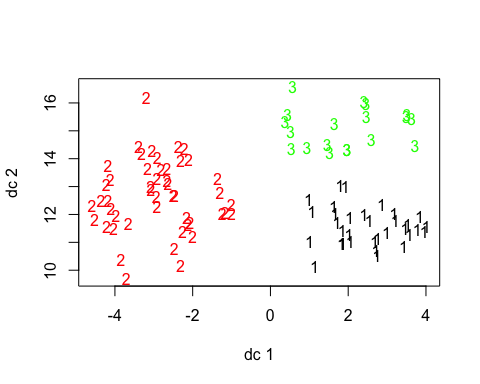
Para ello utilizamos la función princomp()con las bases originales.

Podemos visualizar facilmente la varianza explicada por cada componente principal al utilizar las correlaciones (están acotadas entre -1 y 1). La variación explicada por los componentes principales es igual al número de variables originales y la varianza explicada por cada componentes estará entre 0 y 7. Vemos que con cuatro cuatro componentes explicamos el 90% de la variación. Ahora utilizamos la puntuación de los clientes en los nuevos componentes principales. Esa puntuación está recogida en el objeto scores de la lista bases.acp. La asignamos al objeto bases.puntos.

Si queremos visualizar el resultado de la clasificacion que hemos realizado antes, podemos utilizar la función plot() de esta forma:

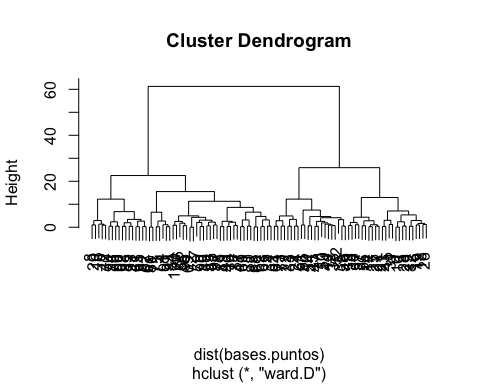
## DELSPEED PRICELEV PRICEFLE MANUFIMA SERVICE SALESFOR  
## Comp.1 0.5279613 -0.7923669 0.6920401 -0.5639719 -0.1858166 -0.4921024  
## Comp.2 0.7515142 0.0930608 0.3744908 0.6020180 0.7789047 0.6039795  
## Comp.3 -0.2024388 -0.5081006 0.1726837 0.4523756 -0.5948887 0.5418084  
## PRODUCTQ  
## Comp.1 -0.738585290  
## Comp.2 -0.269812649  
## Comp.3 0.005434873



Y si queremos visualizar el resultado de la segmentación, podemos utilizar la función plotcluster() del paquete fpc. 

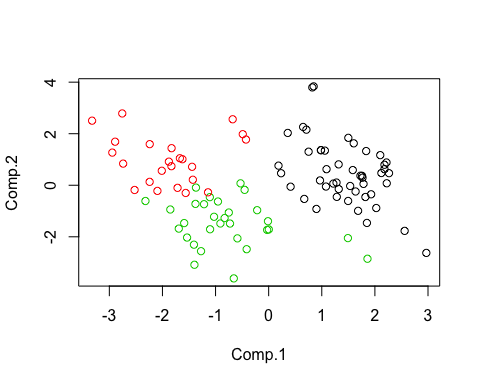
Ahora repetimos el proceso de exploración

## The "ward" method has been renamed to "ward.D"; note new "ward.D2"



## 1 2 3 4 5 6  
## 1 1.2638684 0.4285982 0.06955213 0.03730867 -0.04070992 0.010049399  
## 2 -2.0441421 0.6962137 -0.25451531 -0.29347518 0.16230789 -0.034667748  
## 3 -0.9699069 -1.6025417 0.07681341 0.18507441 -0.05718492 0.009167622  
## 7  
## 1 0.002705582  
## 2 -0.017724773  
## 3 0.010160149

## De nuevo partimos la muestra con kmeans



# Ahora vamos a comprar el resultado actual con el obtenido con las bases originales.

##   
## 1 2 3  
## 1 31 0 2  
## 2 0 20 28  
## 3 16 3 0

## K-means clustering with 3 clusters of sizes 47, 23, 30  
##   
## Cluster means:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## 1 1.4355521 0.4645964 0.00570132 -0.002014796 0.007009984 0.005049449  
## 2 -1.8905322 0.9867716 -0.32424587 -0.094078701 0.024381622 0.095925429  
## 3 -0.7996235 -1.4843925 0.23965643 0.075283518 -0.029674885 -0.081453633  
## Comp.7  
## 1 0.0007120409  
## 2 -0.0251065099  
## 3 0.0181327935  
##   
## Clustering vector:  
## [1] 1 2 3 2 1 3 1 1 3 3 2 2 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 1 3 3 3 1 2  
## [36] 3 3 2 1 2 1 1 1 2 2 1 2 1 2 2 2 2 3 2 1 3 3 2 1 1 2 3 3 2 1 2 1 1 1 1  
## [71] 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 208.11739 90.97915 99.02064  
## (between\_SS / total\_SS = 43.1 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss"   
## [5] "tot.withinss" "betweenss" "size" "iter"   
## [9] "ifault"