

εσταδιστιχ̄

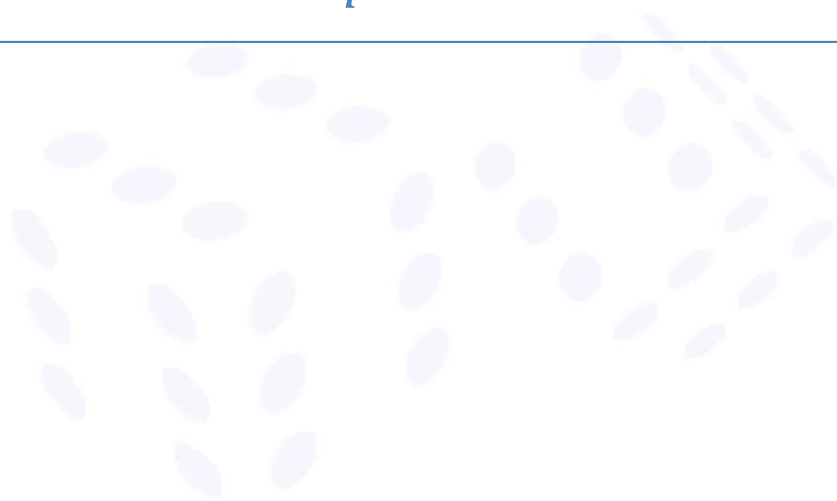
εσταδιστιχ̄

Marketing cuantitativo

Grado en Marketing UOC

ΕΣΤΑΔΙΣΤΙΧ

Apuntes



PEC 1: TEORÍA

SEGMENTACIÓN DE MERCADOS

La **segmentación** debe entenderse como la identificación, en un mercado, de distintos grupos de consumidores que tienen diferentes necesidades, para diseñar estrategias de marketing distintas y atenderlos debidamente. Las empresas deben valorar los segmentos que identifican; para conseguirlo, se fijarán en la competencia existente dentro de ellos, en el tamaño del segmento, en su estabilidad y en su previsible rentabilidad. Pueden seguir estrategias diversas: el **marketing concentrado** se centra en algún segmento y se olvida del resto (por ejemplo, Rolex u Omega) mientras que el **marketing diferenciado** identifica varios segmentos y atiende a todos ellos porque posee recursos financieros y humanos que se lo permiten. En los cuatro casos prácticos del módulo se utilizan las siguientes técnicas de segmentación:

- **Caso I.1:** regresión por árbol de decisión binario (DAI)
- **Caso I.2:** árbol de decisión binario basado en el estadístico Chi cuadrado (CHAID)
- **Caso I.3:** análisis de correspondencias múltiples (ACM) + análisis de clasificación o clúster (AC)
- **Caso I.4:** análisis de correspondencias múltiples (ACM) + análisis de clasificación o clúster (AC)

Técnicas de dependencia: aquellas que permiten explicar la variable dependiente en función de un conjunto de variables independientes explicativas. Su objetivo es determinar si el conjunto de variables independientes afecta y de qué manera a la variable dependiente. Lo son el DAI y CHAID.

Técnicas de interdependencia: aquellas que permiten explicar la relación existente entre todas las variables independientes implicadas. Lo son el ACS y ACM.

Los **criterios principales** de segmentación pueden ser **objetivos** si son características intrínsecas del individuo, ya sean geográficos, demográficos y sociodemográficos, y **subjetivos**, vinculados a su comportamiento, pudiendo ser:

- **Ocasión:** si un producto se consume solo en ciertas circunstancias. Ejemplo: café soluble en sobre en una cafetería o procedente de un frasco en casa.
- **Beneficios buscados:** si compran los productos por determinados atributos que encuentran en ellos. Por ejemplo, busco coche diésel porque es más duradero o más económico en su consumo.
- **Uso:** si consumen determinados bienes o servicios. Por ejemplo, si bebe vino o no.
- **Nivel o intensidad de uso:** si difieren por el grado o frecuencia de consumo. Por ejemplo, hay personas que consumen diez cafés al día y otras solo con el desayuno.
- **Fidelidad:** si consumen siempre la misma marca o acuden al mismo establecimiento para sus compras.
- **Actitudes:** si su actitud es hostil, indiferente o favorable.
- **Situación:** si los bienes o servicios se compran para resolver una necesidad sin alternativa. Por ejemplo, la compra de una cámara de un solo uso, desechable, para sacar unas fotos en un viaje improvisado.

La técnica de segmentación utilizada, muchas veces depende del **tipo de variables** que tenemos:

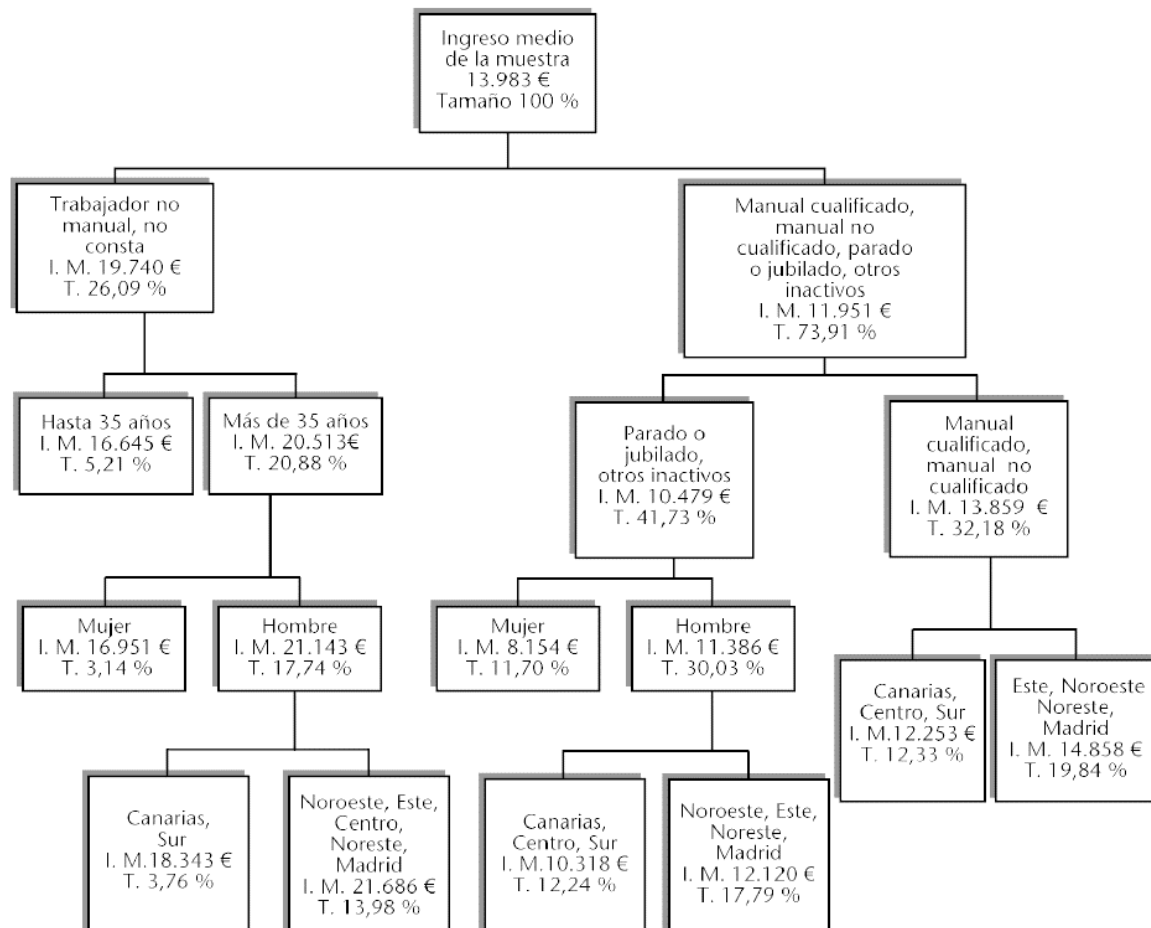
- **Variables cualitativas:** si los valores son categorías (dicotómica/binaria o politómica/categorica).
- **Variables cuantitativas:** si los valores son números (escala/valorativa o continua/métrica).

SEGMENTACIÓN POR REGRESIÓN POR ÁRBOL DE DECISIÓN BINARIA O DETECTOR AUTOMÁTICO DE INTERACCIONES (DAI/AID)

La **regresión por árbol de decisión binaria o detector automático de interacciones (DAI)**, es una técnica de **dependencia** donde la **variable dependiente** de interés es **cuantitativa**, mientras que las **variables independientes** o explicativas son **cualitativas** y me servirán para **segmentar el mercado**.

Se buscará que los grupos o segmentos sean homogéneos internamente y heterogéneos entre ellos.

Usaremos el **análisis de la varianza (ANOVA)**, para detectar las diferencias entre los grupos. Si el p-valor es menor 0,05, quiere decir que la diferencia es significativa.



Con R Commander:

Estadísticos / Medias / ANOVA de un factor

Datos / Conjunto de datos activo / Filtrar el conjunto de datos activo

```

> AnovaModel.1 <- aov(Notas_final ~ Academia, data=Dataset)
> summary(AnovaModel.1)
          Df Sum Sq Mean Sq F value    Pr(>F)
Academia   1  88.84   88.84  23.67 0.0000274 ***
Residuals 33 123.85    3.75

> with(Dataset, numSummary(Notas_final, groups=Academia, statistics=c("mean", "sd")))
      mean      sd data:n
No 4.318182 1.936771    22
Si 7.615385 1.938146    13

> Academiarsi <- subset(Dataset, subset=Academia=="Si")
...
  
```

Ejemplo DAI: en un estudio sobre un determinado buscador de Internet, se llevó a cabo una encuesta electrónica entre los clientes del servicio que da este buscador para que diesen su valoración sobre la utilidad que percibían con su uso. La encuesta fue contestada por 200 usuarios de este buscador. A estos usuarios, aparte de pedirles que lo valorasen de 1 a 10, de menos a más valorado, se les pedía que dijeran la edad y el lugar de conexión con la codificación de la tabla. Se pide hacer una detección automática de interacciones para configurar grupos homogéneos que tengan una valoración parecida sobre el buscador objeto de estudio.

Primero tenemos que modificar las variables en

Datos / Modificar variables del conjunto de datos activo / Recodificar variables

Edat.1: De 18 a 30 - Más de 30 Edat.2: De 31 a 45 - De 18 a 30 y más de 45

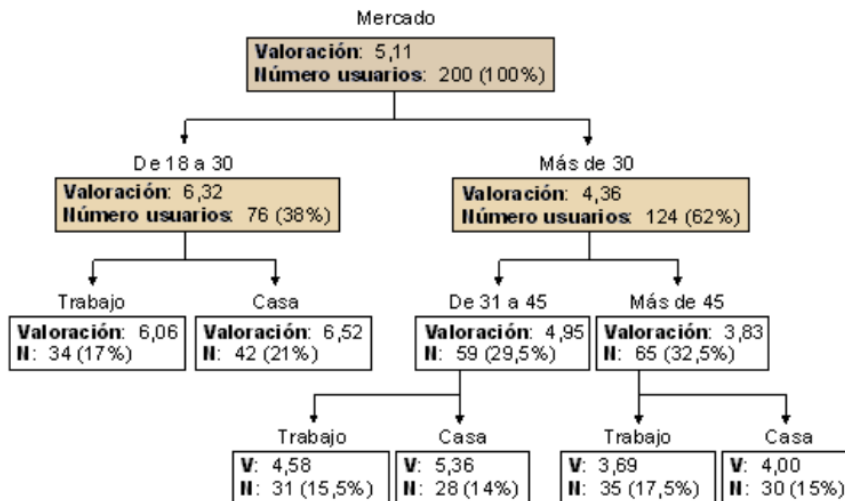
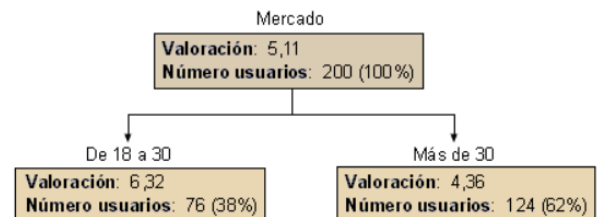
Edat.3: Más de 45 - Menos de 45 Lloc: Trabajo - Casa

Edad	Código
De 18 a 30	1
De 31 a 45	2
Más de 45	3
Lugar de conexión	Código
Trabajo	1
Casa	2

```
> summary(AnovaModel.1)
          Df Sum Sq Mean Sq F value Pr(>F)
Edat.1     1  179.7   179.70   24.12 1.89e-06 ***
Residuals 198 1475.1     7.45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

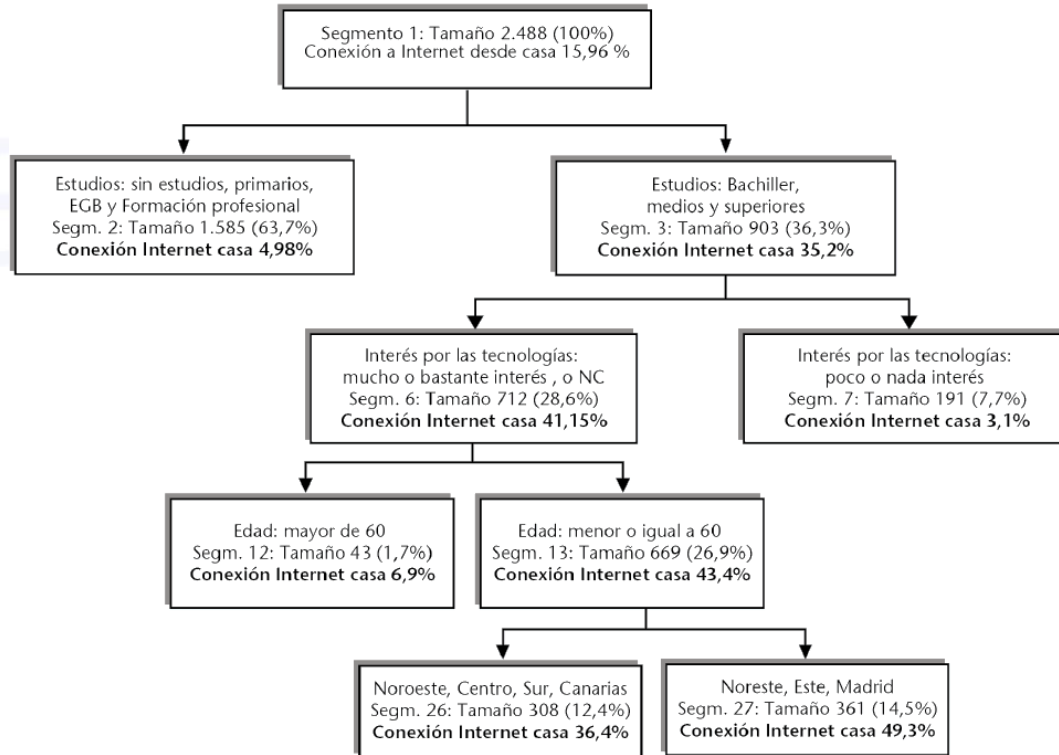
> numSummary(Datos$Valor , groups=Datos$Edat.1, statistics=c("mean", "sd"))
      mean      sd data:n
0 6.315789 2.634188     76
1 4.362903 2.785954    124
```

Dicotomías	F	Grupos	N	Valoración media
Edat.1	24,12	De 18 a 30	76	6,32
		Más de 30	124	4,36
Edat.2	0,24	De 31 a 45	59	4,95
		De 18 a 30 y Más de 45	141	5,17
Edat.3	20,66	Más de 45	65	3,83
		Menos de 45	135	5,72
Lloc.factor	2,72	Trabajo	100	4,77
		Casa	100	5,44



ÁRBOL DE DECISIÓN BINARIA BASADO EN EL ESTADÍSTICO χ^2 (CHAID)

El **árbol de decisión binaria basado en el estadístico Chi Cuadrado (CHAID)**, es una técnica de **dependencia** donde la **variable dependiente** de interés es **cualitativa**, igual que las **variables independientes** o explicativas que también serán **cualitativas** y me servirán para **segmentar el mercado**. A diferencia del DAI, donde trabajábamos con pruebas ANOVA, ahora identificaremos las diferencias con el **estadístico Chi-Cuadrado**.



Con R Commander:

Estadísticos / Tablas de contingencia / Tabla de doble entrada

Frequency table:

		Academia	
Sexo	No	Si	
Hombre	11	4	
Mujer	11	9	

Expected counts:

		Academia	
Sexo	No	Si	
Hombre	9.428571	5.571429	
Mujer	12.571429	7.428571	

Pearson's Chi-squared test

data: .Table

X-squared = 1.234, df = 1, p-value = 0.2666

...

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES (ACM)

El **análisis de correspondencias múltiples (ACM)** es una técnica de **interdependencia** donde únicamente tengo una serie de **variables independientes cualitativas** de las que me interesa relacionar sus diferentes valores. Si solo tengo dos variables cualitativas pasa a ser un **análisis factorial de correspondencias simples (ACS)**. Lo más interesante del análisis es el **gráfico de posicionamiento** donde podremos agrupar las respuestas de las variables por proximidad y así estudiar sus relaciones o asociaciones. Las variables del estudio pueden ser **activas**, si son de interés para el posicionamiento, o **ilustrativas**, si son de caracterización.

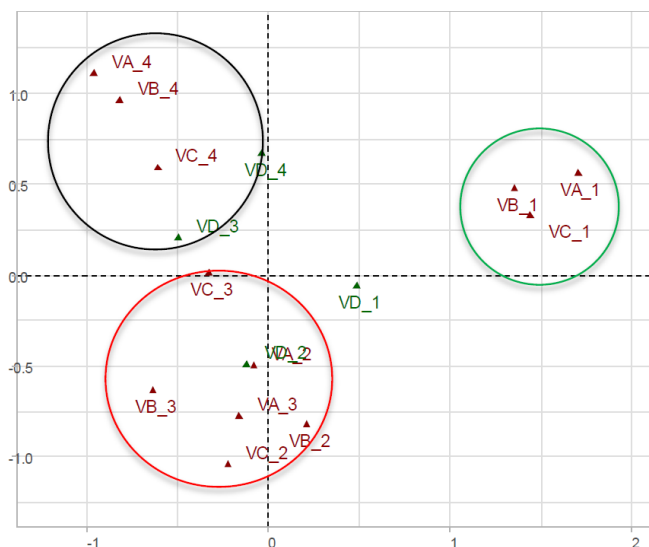
El análisis de correspondencias es una técnica de **análisis factorial**, ya que tiene como objetivo reducir el número de variables con las que trabajamos a través de agrupaciones en dimensiones. Como el análisis agrupa valores de las variables y no a los individuos, si lo que quiero es segmentar individuos o consumidores, después del ACM debería hacer un **análisis de clasificación o clúster** con sus puntuaciones factoriales. Este análisis lo veremos a continuación. Para hacer un análisis de correspondencias, necesitaremos el paquete **FactoMineR** instalado y cargado.

Con R Commander:

FactoMineR / Multiple Correspondence Analysis

Temática (VA)	Miedo/terror VA_1	Misterio/ intriga VA_2	Fantasía VA_3	Ciencia ficción VA_4
Beneficio esperado (VB)	Evasión VB_1	Entretenimiento VB_2	Creatividad VB_3	Disfrute VB_4
Con quién acudiría (VC)	Solo VC_1	Compañeros trabajo VC_2	Amigos VC_3	Familia VC_4

Cohorte generacional (VD)	Baby Boomers VD_1	Generación X VD_2	Millennials VD_3	Centennials VD_4
---------------------------	----------------------	----------------------	---------------------	---------------------



Ejemplo ACM: en un estudio de mercado sobre empresas de distribución de diferente tamaño, en zonas urbanas y rurales, queremos analizar la relación que puede haber entre tres variables categóricas o atributos: el precio de los productos que venden, la variedad de productos y la rapidez en la entrega. Cada una de estas variables tiene las categorías siguientes:

El precio del producto: Alto, Medio y Bajo.

La variedad del producto: Mucha, Normal y Poca.

La rapidez en la entrega: Rápida y Lenta.

Para desarrollar el análisis, se ha pedido a 30 usuarios de estas empresas de distribución que asocien su empresa de compra habitual con cada uno de los atributos. Para analizar la relación entre los atributos, se pide un análisis de correspondencias múltiple. También se pide complementar el análisis con el tamaño de la empresa y la localización.

> res\$eig

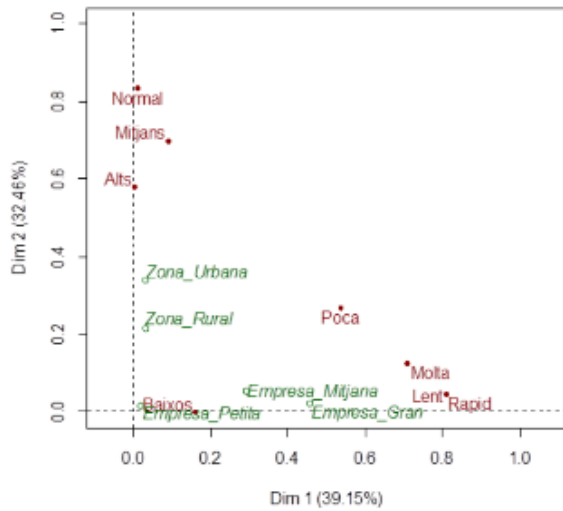
	Eigenvalue	Cumulative percentage of variance	Percentage of variance
dim 1	3.915166e-01	3.915166e+01	39.15166
dim 2	3.246485e-01	3.246485e+01	71.61651
dim 3	1.712577e-01	1.712577e+01	88.74228
dim 4	7.982753e-02	7.982753e+00	96.72503
dim 5	3.274970e-02	3.274970e+00	100.00000
dim 6	3.795874e-32	3.795874e-30	100.00000
dim 7	2.002218e-32	2.002218e-30	100.00000
dim 8	7.327599e-33	7.327599e-31	100.00000

> res\$var

	\$contrib	
	Dim 1	Dim 2
Alts_N	0.06343884	1.188612e+01
Alts_S	0.05550898	1.040035e+01
Mitjans_N	2.03955829	1.881819e+01
Mitjans_S	0.87409641	8.064939e+00
Baixos_N	3.90278721	1.861431e-03
Baixos_S	1.18780480	5.665226e-04
Molta_N	18.89463626	4.052269e+00
Molta_S	3.77892725	8.104537e-01
Normal_N	0.27987566	2.459597e+01
Normal_S	0.08517955	7.485731e+00
Poca_N	6.86897311	4.128181e+00
Poca_S	10.30345966	6.192271e+00
Rapid_N	18.94410979	1.306467e+00
Rapid_S	6.88876720	4.750789e-01
Lent_N	6.88876720	4.750789e-01
Lent_S	18.94410979	1.306467e+00

> res\$quali.sup

	\$cos2	
	Dim 1	Dim 2
Empresa_Gran_N	0.45652235	0.02197919
Empresa_Gran_S	0.45652235	0.02197919
Empresa_Mitjana_N	0.29220774	0.05383722
Empresa_Mitjana_S	0.29220774	0.05383722
Empresa_Petita_N	0.02063508	0.01820710
Empresa_Petita_S	0.02063508	0.01820710
Zona_Urbana_N	0.03321034	0.33977984
Zona_Urbana_S	0.03321034	0.33977984
Zona_Rural_N	0.03168186	0.21634851
Zona_Rural_S	0.03168186	0.21634851



ANÁLISIS DE CLASIFICACIÓN, CONGLOMERADOS O CLUSTER

El **análisis de clasificación, de conglomerados o clúster (AC)** es una técnica de **interdependencia** donde únicamente tengo una serie de **variables independientes cuantitativas** de las que me interesa relacionar sus diferentes valores para segmentar a los individuos. El **análisis jerárquico** nos permitirá agrupar a los individuos sin un número de grupos definido a partir de la representación gráfica, el **dendrograma**. En el caso del **análisis no jerárquico**, el número de **grupos se definen** previamente. El AC se puede usar como complemento del análisis de correspondencias simples y múltiples (ACS y ACM).

Con R Commander:

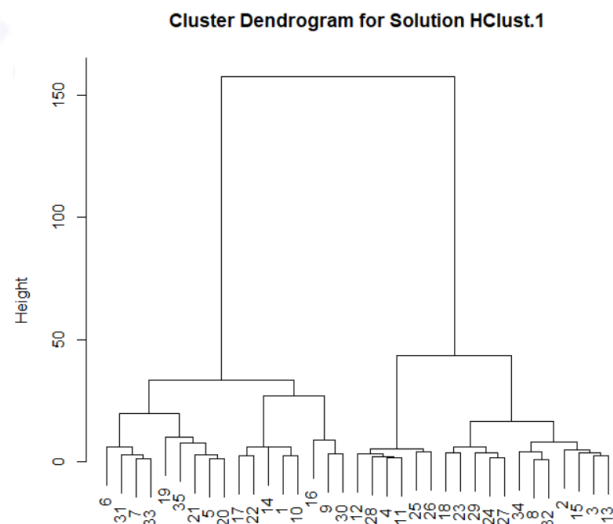
Estadísticos / Análisis dimensional / Análisis de agrupación / Agrupación jerárquica

Estadísticos / Análisis dimensional / Análisis de agrupación / Resumir agrupación jerárquica

Estadísticos / Análisis dimensional / Análisis de agrupación / Agregar agrupación jerárquica al conjunto de datos

Si quiero una agrupación no jerárquica (si tengo el número de grupos previamente definido)

Estadísticos / Análisis dimensional / Análisis de agrupación / Agrupación por k-medias



```
> summary(as.factor(cutree(HClust.1, k = 4))) # Cluster Sizes
 1  2  3  4
 8 12  6  9
```

```
> by(model.matrix(~-1 + Asistencia + Edad + Notas_final + Valoracion_final, Dataset),
+ as.factor(cutree(HClust.1, k = 4)), colMeans) # Cluster Centroids
```

INDICES: 1

Asistencia	Edad	Notas_final	Valoracion_final
10.625	25.125	3.250	1.250

INDICES: 2

Asistencia	Edad	Notas_final	Valoracion_final
19.916667	20.083333	6.083333	2.916667

INDICES: 3

Asistencia	Edad	Notas_final	Valoracion_final
27.666667	20.166667	8.833333	4.666667

INDICES: 4

Asistencia	Edad	Notas_final	Valoracion_final
5.222222	20.555556	4.666667	3.444444

Ejemplo AC: Disponemos de datos de algunos países, como Alemania, Bélgica, España, Estados Unidos, Finlandia, Francia, Noruega, Portugal y el Reino Unido referentes a aspectos relacionados con el desarrollo de la sociedad del conocimiento para el año 2008:

A1: usuarios de Internet (por cada 100 habitantes).

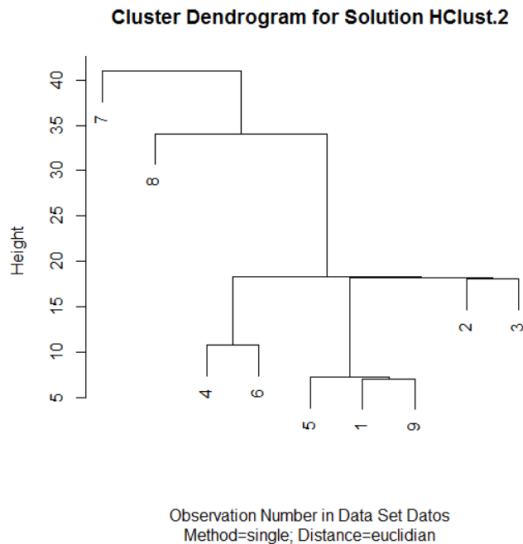
A2: usuarios de banda ancha (por cada 100 habitantes).

A3: usuarios de teléfonos móviles (por cada 100 habitantes).

A4: nivel de PIB por habitante (en dólares del 2007).

	A1	A2	A3	A4
Alemania	75,33	27,47	128,27	40,308
Bélgica	68,86	27,97	111,63	43,565
España	56,74	20,22	111,67	32,647
Estados Unidos	74,00	23,46	86,79	44,841
Finlandia	82,62	30,50	128,76	39,828
Francia	68,21	28,52	93,45	41,463
Noruega	82,55	33,27	110,16	82,716
Portugal	41,92	15,31	139,64	20,965
Reino Unido	79,62	28,21	126,34	45,510

1.- Hacer un análisis jerárquico:



Los países se agrupan de la siguiente manera:

1^{er} grupo (agrupación 1): Alemania (1), Finlandia (5), el Reino Unido (9), Bélgica (2), España (3), los Estados Unidos (4) y Francia (6).

2^o grupo (agrupación 3): Portugal (8).

3^{er} grupo (agrupación 2): Noruega (7).

```
> summary(as.factor(cutree(HClust.2, k = 3))) # Cluster Sizes
1 2 3
7 1 1
```

```
> by(model.matrix(~-1 + A1 + A2 + A3 + A4, Datos), as.factor(cutree(HClust.2, k = 3)), colMeans) # Cluster Centroids
INDICES: 1
      A1      A2      A3      A4
72.19714 26.62143 112.41571 41.16600
-----
INDICES: 2
      A1      A2      A3      A4
82.550  33.270 110.160  82.716
-----
INDICES: 3
      A1      A2      A3      A4
41.920  15.310 139.640  20.965
```

2.-Hacer un análisis no jerárquico con 4 grupos:

```
> .cluster$size # Cluster Sizes
[1] 5 2 1 1
> .cluster$centers # Cluster Centroids
new.x.A2 new.x.A3 new.x.A4
1  26.874 121.334 40.3716
2  25.990  90.120 43.1520
3  33.270 110.160 82.7160
4  15.310 139.640 20.9650
> .cluster$withinss # Within Cluster Sum of Squares
[1] 473.20899 40.68504 0.00000 0.00000
> .cluster$tot.withinss # Total Within Sum of Squares
[1] 513.894
> .cluster$betweenss # Between Cluster Sum of Squares
[1] 4334.594
```


PEC1. Segmentación de Mercados.

b) ¿Se podría haber aplicado un análisis de regresión por árbol de decisión binario, ya que son técnicas similares y se interpretan de la misma manera?

3. (1 punto) Una plataforma audiovisual quiere segmentar a sus clientes en función del uso de la misma, siendo la variable los minutos consumidos. Para hacer dicha segmentación han obtenido de su base de datos información relevante de cada uno de los usuarios. Las variables son las siguientes:













- Principales días de consumo: (a) entre semana, (b) fin de semana, (c) Indistinto
- Principales momentos del día de consumo: (a) Mañanas, (b) Tardes, (a) Noches
- Formatos de consumo: (a) Películas, (b) Series, (c) Documentales.
- Género: (a) Home; (b) Mujer; (c) No definido.
- Edad: a) <20 años; b) 20-40 años; c) 41-60 años; d) >60 años.

Con esta información contestad a las siguientes preguntas **justificando vuestra respuesta (siete líneas como máximo por apartado)**:

a) ¿Qué criterios de segmentación se realiza según la naturaleza de las variables?

b) Según el objetivo marcado en el enunciado, ¿qué técnica se debería de aplicar?

4. (1 punto) un teatro pequeño de una determinada ciudad quiere conocer la opinión de sus espectadores y saber qué aspectos influyen a la hora de volver a venir a ver otra obra o espectáculo. Con este objetivo han dejado en las buracas unos cuestionarios muy sencillos para que los cumplimenten al finalizar la obra y los dejen en la entrada. Se les preguntaba lo siguiente:

Calidad de la obra (AT1)	  
Comodidad de la sala (AT2)	  
Servicios del bar (AT3)	  
Adecuación de días/horas del espectáculo (AT4)	  
¿Volverás a escogernos para ver otra obra o espectáculo? (PR)	Si No

Con esta información recopilada han realizado el siguiente análisis:

PEC1. Segmentación de Mercados.

```

Call:
"res<-MCA(Dataset.MCA, ncp=5, graph = FALSE)"
> res$eig
  eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.46673050                25.929472                25.92947
dim 2 0.35533544                19.740858                45.67033
dim 3 0.24883466                13.824148                59.49448
dim 4 0.19653074                10.918375                70.41285
dim 5 0.16157859                8.976589                 79.38944
dim 6 0.13147154                7.303975                 86.69342
dim 7 0.09591277                5.328487                 92.02190
dim 8 0.07794275                4.330153                 96.35206
dim 9 0.06566300                3.647944                 100.00000

> res$var
$coord
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
AT1_ AMARILLO  0.6084696 -1.01339542 -0.04877874 -0.25364563 -0.19779314
AT1_ ROJO      1.1708746  1.37889053  0.25642651 -0.73530746  1.25209851
AT1_ VERDE    -0.6554513  0.29903275 -0.03085757  0.33209193 -0.17567532
AT2_ AMARILLO  0.2884674  -0.79883993  0.54413247  0.38837245  0.63266339
AT2_ ROJO      0.9156847  1.23196628  0.25168237  0.75716761 -1.06838279
AT2_ VERDE    -0.5347300  0.09269237 -0.46502625 -0.54468501 -0.03926024
AT3_ AMARILLO  0.3766142  -0.84933109  0.27882990  0.29323561 -0.14187700
AT3_ ROJO      2.3303663  1.55657095  0.03807339 -0.07751464  0.91632774
AT3_ VERDE    -0.6019196  0.38429779 -0.20460327 -0.19838049 -0.02956325
AT4_ AMARILLO  0.4617055  0.10272654  0.84367262 -0.69502960 -0.49825354
AT4_ ROJO      0.3123970  -0.45168331 -1.25718790 -0.05961241  0.12028668
AT4_ VERDE    -0.7855878  0.30630985  0.25191062  0.80352026  0.42554688
NO             1.0059442  0.15453717 -0.63794644  0.42508609 -0.15139342
SI            -0.5029721 -0.07726858  0.31897322 -0.21254304  0.07569671

$contrib
      Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
AT1_ AMARILLO  5.288351 19.2676206  0.063746885  2.18239330  1.61416288
AT1_ ROJO      7.531647 13.7200737  0.677564044  7.05412000 24.87875039
AT1_ VERDE     9.912868  2.7100897  0.041209489  6.04325376  2.05694353
AT2_ AMARILLO  1.188601 11.9726349  7.932446289  5.11652470 16.51468610
AT2_ ROJO      6.448952 15.3328379  0.913814704 10.47169508 25.35907267
AT2_ VERDE     5.969279  0.2355963  8.467652729 14.70887086  0.09294822
AT3_ AMARILLO  2.337673 15.6160851  2.403391096  3.36557721  0.95828968
AT3_ ROJO     17.900652 10.4902529  0.008962287  0.04703526  7.99474252
AT3_ VERDE     8.359792  4.4759160  1.811752786  2.15651301  0.05825124
AT4_ AMARILLO  3.279121  0.2132162 20.536690150 17.64694480 11.03088584
AT4_ ROJO      1.286750  3.5332648 39.087341774  0.11127293  0.55105872
AT4_ VERDE     8.815198  1.7603221  1.700164049 21.90140130  7.47170540
NO             14.454077  0.4480599 10.903509146  6.12959852  0.94566854

```


PEC1. Segmentación de Mercados.

VARIABLE A EVALUAR		ESCALA DE MEDIDA
P1	Esta bodega ofrece productos de buena calidad	1: muy en desacuerdo...7: muy de acuerdo (escala Likert de 7 puntos)
P2	He disfrutado visitando las instalaciones de la bodega	
P3	Comprar vinos de esta bodega es una buena opción	
P4	Durante mi visita, han conseguido que me sienta satisfecho	
P5	La bodega me ha informado sobre sus productos de manera muy detallada	

Además, han preguntado sobre el motivo de la visita y los siguientes datos de clasificación:

GÉNERO (CLA1)	MOTIVO DE LA VISITA (CLA2)	EDAD (CLA3)
HOMBRE	OCIO	<= 25 AÑOS
MUJER	FAMILIAR	26-40 AÑOS
	NEGOCIOS	>40 AÑOS

Encontraréis la base de datos en el archivo 78585_20231_PEC1_EJ6.xlsx.

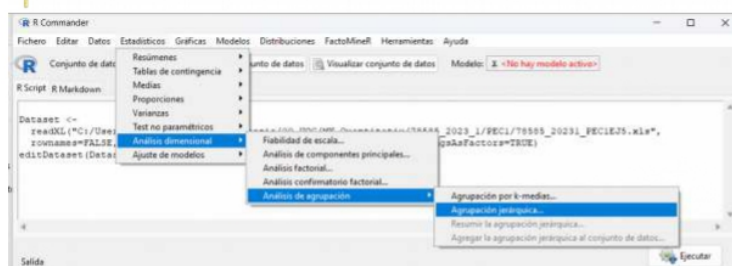
A partir de la información anterior, contestad las siguientes preguntas, **argumentando vuestra respuesta en cada apartado**:

- a) (0,75 puntos) Ejecutad con RCommander un análisis clúster para conseguir segmentar a los usuarios a los clientes de la bodega. Escoged únicamente las variables adecuadas para esta técnica. Justificad cada una de vuestras decisiones.

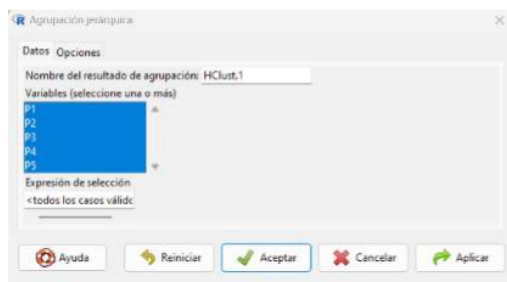
En este caso, podríamos utilizar tanto el análisis de conglomerados jerárquico como el k-medias, ya que tenemos un nº considerable de clientes a segmentar (135 en total). Como no sabemos inicialmente cuantos segmentos debemos extraer podemos realizar un clúster jerárquico para identificar, a través del dendograma, cuál sería el número ideal de segmentos a analizar.

Las variables que utilizaremos para este análisis son las preguntas que van de la P1 a P5, ya que son cuantitativas y válidas para este análisis. Inicialmente no haría falta que estandarizemos los datos ya que tienen la misma escala.

Lo primero que haremos es aplicar el análisis de conglomerados jerárquico para que nos muestre el dendograma y podamos valorar en cuantos grupos podemos segmentar la muestra. Seguiremos la siguiente ruta: *Estadísticos >> Análisis dimensional >> Análisis de agrupación >> Agrupación jerárquica*:

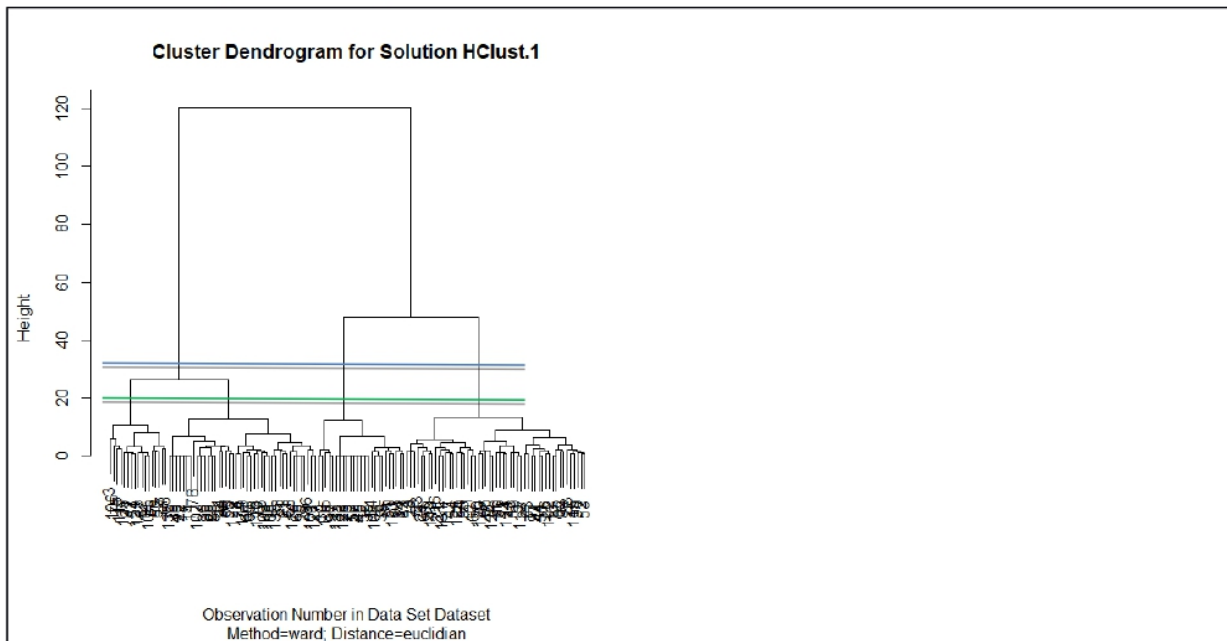


Seleccionamos las variables para realizar el análisis:



PEC1. Segmentación de Mercados.

Y obtenemos el siguiente dendograma:



Vemos a través del dendograma que un número apropiado de segmentos podrían ser 3 segmentos (línea verde) o 4 segmentos (línea azul). Obtendremos la salida de resultados con 3 y 4 segmentos y a la hora de interpretarlo decidiremos cuál de las dos puede ser más útil.

3 segmentos:

```
> summary(as.factor(cutree(HClust.1, k = 3))) # Cluster Sizes
 1  2  3
59 51 25

> by(model.matrix(~-1 + P1 + P2 + P3 + P4 + P5, Dataset),
as.factor(cutree(HClust.1, k = 3)), colMeans)
+ # Cluster Centroids
INDICES: 1
      P1      P2      P3      P4      P5
4.288136 4.593220 3.932203 4.084746 4.491525
-----
INDICES: 2
      P1      P2      P3      P4      P5
4.980392 5.725490 5.607843 5.588235 5.901961
-----
INDICES: 3
      P1  P2  P3  P4  P5
6.20 6.84 6.88 6.88 6.88

> biplot(princomp(model.matrix(~-1 + P1 + P2 + P3 + P4 + P5, Dataset)), xlabs
= as.character(cutree(HClust.1, k =
+ 3)))
```

4 segmentos:

```
> summary(as.factor(cutree(HClust.1, k = 4))) # Cluster Sizes
```


PEC1. Segmentación de Mercados.

```

1 2 3 4
42 51 17 25

> by(model.matrix(~-1 + P1 + P2 + P3 + P4 + P5, Dataset),
as.factor(cutree(HClust.1, k = 4)), colMeans)
+ # Cluster Centroids
INDICES: 1
      P1      P2      P3      P4      P5
4.642857 4.928571 4.214286 4.452381 4.785714
-----
INDICES: 2
      P1      P2      P3      P4      P5
4.980392 5.725490 5.607843 5.588235 5.901961
-----
INDICES: 3
      P1      P2      P3      P4      P5
3.411765 3.764706 3.235294 3.176471 3.764706
-----
INDICES: 4
      P1  P2  P3  P4  P5
6.20 6.84 6.88 6.88 6.88

> biplot(princomp(model.matrix(~-1 + P1 + P2 + P3 + P4 + P5, Dataset)), xlabs
= as.character(cutree(HClust.1, k =
+ 4)))

```

- b) (1,25 puntos) Interpretad los resultados detalladamente (nº de segmentos a retener, descripción de los segmentos, etc.)

PEC1. Segmentación de Mercados.

- c) (0,75 puntos) Quieren saber si existen diferencias en la distribución de los segmentos según las variables género, motivo de la visita y edad. Aplicad la técnica adecuada para dar respuesta a esta necesidad e interpretad los resultados.

```
> local({
+   .Table <- xtabs(~hclus.label+CLA1, data=Dataset)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   cat("\nRow percentages:\n")
+   print(rowPercents(.Table))
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+ })
```

Frequency table:

	CLA1	
hclus.label	HOMBRE	MUJER
1	30	12
2	29	22
3	13	4
4	14	11

Row percentages:

	CLA1		Total	Count
hclus.label	HOMBRE	MUJER		
1	71.4	28.6	100	42
2	56.9	43.1	100	51
3	76.5	23.5	100	17
4	56.0	44.0	100	25

Pearson's Chi-squared test

data: .Table

X-squared = 3.9562, df = 3, p-value = 0.2662

```
> local({
+   .Table <- xtabs(~hclus.label+CLA2, data=Dataset)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   cat("\nRow percentages:\n")
+   print(rowPercents(.Table))
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+ })
```

Frequency table:

PEC1. Segmentación de Mercados.

```

          CLA2
hclus.label FAMILIAR NEGOCIOS OCIO
      1         7         14    21
      2         9         13    29
      3         4         11     2
      4         3          6    16

Row percentages:
          CLA2
hclus.label FAMILIAR NEGOCIOS OCIO Total Count
      1      16.7      33.3 50.0    100    42
      2      17.6      25.5 56.9    100    51
      3      23.5      64.7 11.8    100    17
      4      12.0      24.0 64.0    100    25

Pearson's Chi-squared test

data: .Table
X-squared = 13.929, df = 6, p-value = 0.03044

> local({
+   .Table <- xtabs(~hclus.label+CLA3, data=Dataset)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   cat("\nRow percentages:\n")
+   print(rowPercents(.Table))
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+ })

Frequency table:
          CLA3
hclus.label <= 25 AÑOS >40 AÑOS 26-40 AÑOS
      1         4         21         17
      2         6         19         26
      3         0          8          9
      4         5          8         12

Row percentages:
          CLA3
hclus.label <= 25 AÑOS >40 AÑOS 26-40 AÑOS Total Count
      1         9.5      50.0      40.5 100.0    42
      2        11.8      37.3      51.0 100.1    51
      3         0.0      47.1      52.9 100.0    17
      4        20.0      32.0      48.0 100.0    25

Pearson's Chi-squared test

data: .Table
X-squared = 6.0798, df = 6, p-value = 0.4143

```

PEC1. Segmentación de Mercados.

d) (0,75 puntos) Según toda la información obtenida, ¿le recomendarías a los responsables de la bodega focalizarse en mejorar la satisfacción de algún segmento en concreto? Argumentad vuestra respuesta.

5. (1 punto) La federación de Alpinismo se plantea realizar una campaña a de formación para los socios federados y poder mejorar así su experiencia en la montaña. Para desarrollar las diferentes posibles formaciones han decidido primero conocer un poco más al abonado mediante un pequeño cuestionario que han enviado online. En este cuestionario se les pregunta por los motivos de sus excursiones, la duración y la formación que tienen previamente. A continuación, se detallan las preguntas y alternativas de respuestas;

AT1: Motivaciones/motivo principal
AT1_1: Diversión/pasarlo bien
AT1_2: Superación personal/ asumir objetivos cada vez más difíciles
AT1_3: Compartir buenos momentos con amigos
AT1_4: Desconectar del trabajo/familia/obligaciones
AT2: Duración máxima de la salida/expedición:
AT2_1: Un día
AT2_2: Un fin de semana
AT2_3: Una semana
AT2_4: Más de una semana
AT3: Tipo de formación:
AT3_1: Cursos de escuela de formación
AT3_2: Libros
AT3_3: A través de guías de montaña
AT3_4: A través de amigos con mayor experiencia
AT3_5: A través de la práctica

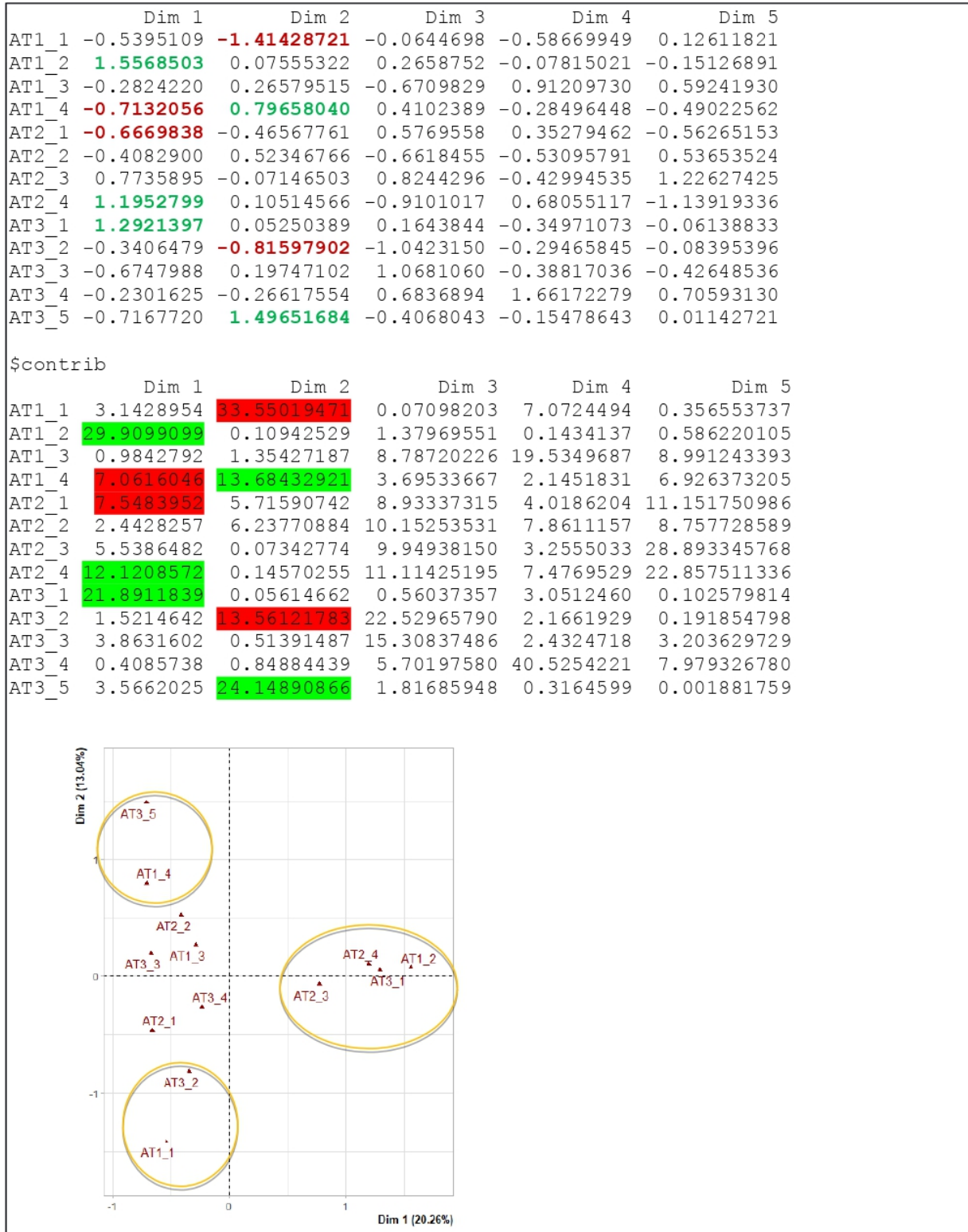
Tras la finalización del trabajo de campo, han realizado el siguiente análisis:

```
Call:
"res<-MCA(Dataset.MCA, ncp=5, graph = FALSE)"

> res$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1  0.67530098           20.259029           20.25903
dim 2  0.43471737           13.041521           33.30055
dim 3  0.42696379           12.808914           46.10946
dim 4  0.35488560           10.646568           56.75603
dim 5  0.32528003            9.758401           66.51443
dim 6  0.32186233            9.655870           76.17030
dim 7  0.27353615            8.206085           84.37639
dim 8  0.26194545            7.858364           92.23475
dim 9  0.17399693            5.219908           97.45466
dim 10 0.08484471            2.545341          100.00000

> res$var
$coord
```

PEC1. Segmentación de Mercados.



Según toda la información anterior, contestad las siguientes preguntas, **argumentando vuestra respuesta en cada apartado:**

- a) (0,5 puntos) Identificad qué análisis se han aplicado. ¿Se han aplicado de forma correcta? ¿Podríamos haber aplicado un segundo análisis de segmentación para completar la información? Argumentad la respuesta.

PEC1. Segmentación de Mercados.

b) (1,25 puntos) Interpretad los resultados detalladamente (número de dimensiones, categorías que más definen dichas dimensiones, asociación entre categorías que definen los segmentos, etc.). Es imprescindible que señaléis o escribáis los resultados numéricos en los que os basáis para argumentar vuestra respuesta.

c) (0,75 puntos) Según los resultados, ¿recomendarías a la escuela de negocios que apostara por una estrategia de comunicación general o diferenciada?

Este dossier está hecho para seguir la clase de prueba.

Si te apuntas al curso te enviaremos el dossier entero con todos los temas que faltan, ejercicios y exámenes de años anteriores

Más información en:

www.estadistix.com

**Y si tienes cualquier consulta,
escribenos un whatsapp al 644310902**

