

## ADEI-QUIZ1 2014-2015

Octubre 2014. 1 hora

Apellidos

Nombre

Notas: jueves 5 Noviembre. Revisión de notas: pedir hora por mail.

Se analiza una encuesta relativa a la salud, contestada por 100 individuos. El resumen de las variables disponibles (categóricas y cuantitativas) es el siguiente:

Gender	Health.condition	Age_group	Physical_health	Mental.Health
female:68	health_poor :29	18_25 años : 9	Min. : 0.00	Min. : 4.00
male :32	health_fair :20	26_35 años :21	1st Qu.: 43.75	1st Qu.: 52.00
	health_good :34	36_45 años :19	Median : 72.50	Median : 66.00
	health_very good: 9	46_55 años :12	Mean : 65.55	Mean : 62.84
	health_excellent: 8	56_65 años :20	3rd Qu.: 95.00	3rd Qu.: 76.00
		66_75 años :15	Max. :100.00	Max. :100.00
		76 y más años: 4		
Vitality	Freedom_from_pain			
Min. : 5.00	Min. : 0.00			
1st Qu.: 35.00	1st Qu.: 44.44			
Median : 50.00	Median : 66.67			
Mean : 51.95	Mean : 64.33			
3rd Qu.: 70.00	3rd Qu.:100.00			
Max. :100.00	Max. :100.00			

### 1. ACP

Se ha realizada un ACP, tomando como activas las variables: *Physical\_health*, *Mental.Health* y *Vitality*.

Todas las otras variables se consideran suplementarias. Los valores propios son los siguientes: “

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	1.92	63.99	63.99
comp 2	0.84	27.98	91.98
comp 3	0.24	8.02	100.00

#### 1.1 Se decide conservar 2 ejes. Os parece una elección correcta? Razonen su respuesta. 1 pto

*Aunque el segundo valor propio tenga un valor inferior al valor medio, hay muchos argumentos a favor de conservar y estudiar el segundo eje. El primero es que la oposición puesta de relieve sobre este eje es interpretable (Salud física versus salud mental). Además, la información visualizada sobre dos ejes es mucho más legible. En este caso, los dos ejes, conjuntamente, expresan 92% de la inercia, lo que permite afirmar que los ángulos entre las variables sufren muy poca deformación, lo que hace que el ángulo entre dos variables sobre el plano ofrece una muy buen aproximación de este ángulo en el espacio total.*

#### 1.2 A partir de la representación de los individuos y de las variables sobre el primer plano factorial, asignen a las 4 descripciones siguientes, el identificador correcto de los respondientes, sabiendo que pertenecen a la lista siguiente: “I0070” “I0097”, “I0048”, “I0067”),\_ 2 ptos

> base[which(row.names(base))== "I0067",1:7]

Gender Health.condition Age\_group Physical\_health Mental.Health Vitality Freedom\_from\_pain  
female health\_poor 46\_55 años 100 16 35 44.44

> base[which(row.names(base))== "I0097",1:7]

Gender Health.condition Age\_group Physical\_health Mental.Health Vitality Freedom\_from\_pain  
male health\_good 46\_55 años 100 100 90 100

> base[which(row.names(base))== "I0048",1:7]

Gender Health.condition Age\_group Physical\_health Mental.Health Vitality Freedom\_from\_pain  
female health\_excellent 36\_45 años 10 92 80 100

> base[which(row.names(base))== "I0070",1:7]

Gender Health.condition Age\_group Physical\_health Mental.Health Vitality Freedom\_from\_pain  
male health\_poor 56\_65 años 30 20 15 22.22

### 1.3 Hagan una interpretación de los dos primeros ejes a partir de los gráficos y de la descripción de los ejes

#### 3 ptos

*El primer eje es un eje predominante que conserva el 64% de la inercia. El segundo eje conserva un 28% de la inercia, lo que no es despreciable. Conjuntamente, los dos primeros ejes conservan 92% de la inercia, lo que hace que el resumen proveído por el ACP es cuasi-completo: las proximidades entre los individuos sobre el primer plano refleja de forma muy fiel la proximidad en todo el espacio. Los ángulos entre las variables (y entre los ejes y las variables) son también representados de forma fidedigna.*

#### **Primer eje**

*Desde el punto de vista de las variables*

*Todas las variables activas se sitúan en la parte positiva de la primera componente (las correlaciones entre variables activas y la primera componente son todas positivas). Por lo tanto, se observa “un efecto tamaño”: los diversos aspectos de la salud están muy correlacionados y, así, los individuos tienden a tener o bien valores altos o bien valores altos en todos los aspectos de la salud. La primera componente se puede resumir como “nivel de salud. Noten que “ausencia de dolor” tiene también una correlación positiva con esta primera componente.*

*Desde el punto de vista de los individuos*

*A causa de la dualidad existente entre las coordenadas de los puntos-variables y las de los puntos-individuo, podemos afirmar que los individuos con coordenada elevada sobre el primer eje gozan de buena salud en todos sus aspectos (ver pregunta anterior, I0097) y los con coordenada fuertemente negativa sobre este mismo eje gozan de mala salud en todos sus aspectos (ver pregunta anterior, I0070). Cuando a los con coordenada media, su nivel de salud es medio y puede presentar disparidades según los aspectos.*

*Variables suplementarias*

*La variable cuantitativa suplementaria “ausencia de dolor” está correlacionada positivamente con la primera componente, aunque la correlación tenga sólo un valor medio. Lo que concierne al dolor tiene unas características propias que diferencia esta variable de las otras. Podría ser útil e interesante explorar (en otro estudio) dichas características.*

*El posicionamiento de las categorías de la variable categórica “nivel de salud” sobre el primer eje ratifica la interpretación que hemos hecho de este eje.*

*El posicionamiento de las categorías de la variable categórica “edad\_en\_clase” sobre el primer eje muestra la relación entre edad y salud (lo que sabíamos) pero también que el nivel de salud evoluciona poco hasta los 45 años, y se deteriora de forma marcada a partir de la década (46-55 años), queda bastante estable hasta los 65 años y marca otro salto en la década 66-75 años. Esta es una información más relevante que la simple relación entre edad y nivel de salud.*

#### **Segundo eje**

*Desde el punto de vista de las variables*

*El segundo eje opone “salud física” a “salud mental” y vitalidad que podemos considerar, esta última, como un aspecto de la salud muy ligado a la salud mental.*

*Desde el punto de vista de los individuos*

*Esta oposición indica que, para un nivel de salud dado, ciertos individuos benefician de una mejor salud física (coordenada positiva sobre el segundo eje) o de una mejor salud mental (coordenada negativa sobre el segundo eje).*

*Las variables suplementarias utilizadas no guardan relación con el segundo eje.*

## Clasificación de los individuos

Después del ACP, se han clasificado a los individuos a partir de 2 ejes. Los principales resultados se presentan en la Figura 2.

Comparen y caractericen las clases 2 y 3 a partir de su descripción mediante las categorías y las medias de las variables cuantitativas. Comenten los tests efectuados. **4 ptos**

*Las clases 2 y 3 son las dos clases “centrales”, que engloban individuos con niveles de salud intermedios;*

**a. Caracterización por la posición de las clases sobre el primer plano principal,**

*vemos que la clase 2 reúne individuos con mejor salud mental que física, mientras la clase 3 reúne individuos con mejor salud física que mental. Dicha primera interpretación viene confortada por los tests efectuados.*

*Después, se busca caracterizar a las dos clases por las categorías de las variables categóricas suplementarias y por las variables cuantitativas activas y suplementarias.*

**b. Caracterización por las categorías**

*Todas las categorías listadas en los resultados correspondientes a una clase caracterizan dicha clase, aunque unas por estar sobrerrepresentadas en la clase ( $Mod/Cla > Global$ ) y otras por estar infrarepresentadas en la clase ( $Mod/Cla < Global$ ).*

*El test efectuado para una categoría dada compara estas dos proporciones ( $Mod/Cla$  y  $Global$ ):*

*$H_0$ ) en la población,  $Mod/Cla = Global$*

*$H_1$ ) en la población,  $Mod/Cla \neq Global$*

*Menor es el p-value, mayor es la sobre o infra representación. Que una categoría esté sobre representada/infra-representada se ve porque  $Mod/Cla > Global$  /  $Mod/Cla < Global$ .*

*En este caso, ninguna de las categorías corresponde a un  $p\text{-value} < 0.05$ . Sólo la categoría “nivel de salud=health\_good” corresponde a un  $p\text{-value} < 0.1$ . La clase 3 tiene un ligero superávit de individuos con este nivel de salud.*

**c. Caracterización por las variables cuantitativas**

*El test efectuado corresponde a la comparación de la media en la clase con la media global.*

*$H_0$ ) media poblacional de la variable examinada en la clase = media global poblacional de la variable examinada*

*$H_0$ ) media poblacional de la variable examinada en la clase  $\neq$  media global poblacional de la variable examinada*

*Si el p-value es inferior a 0.05 (o a 0.1, si se es más laxo), se rechaza  $H_0$ .*

*Los resultados de la caracterización por las variables cuantitativas confortan y precisan la lectura de la posición de las clases sobre el primer plano factorial.*

- Los individuos de la clase 2 presentan scores medios en “Salud mental” y “Vitalidad” significativamente superiores a la media de toda la muestra y score medio de “Salud física” muy inferior a la media de toda la muestra.*
- Los individuos de la clase 3 presentan un score medio en “salud física” superior a la media de toda la muestra,*

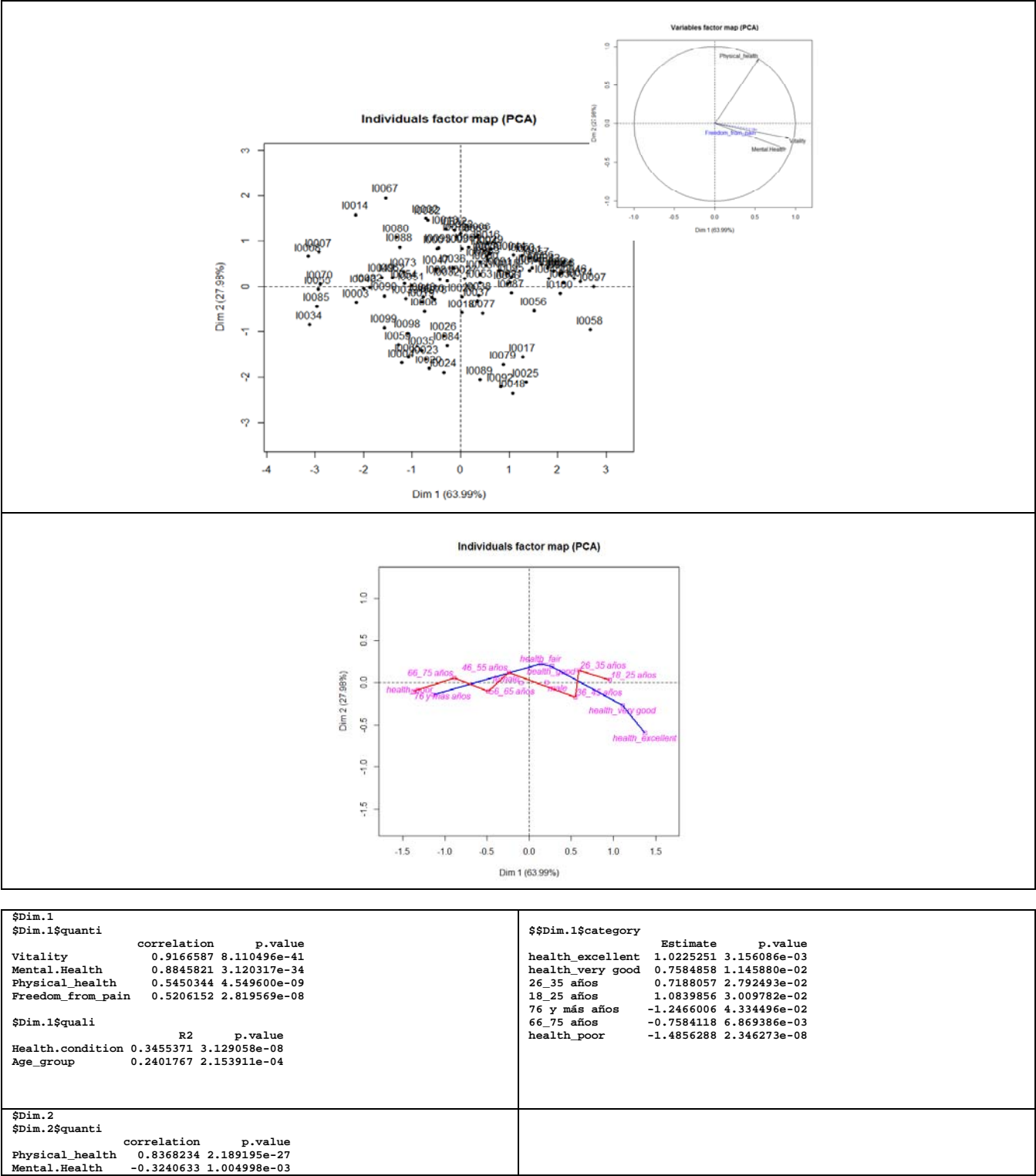


Figura 1. Principales resultados gráficos y numéricos del ACP

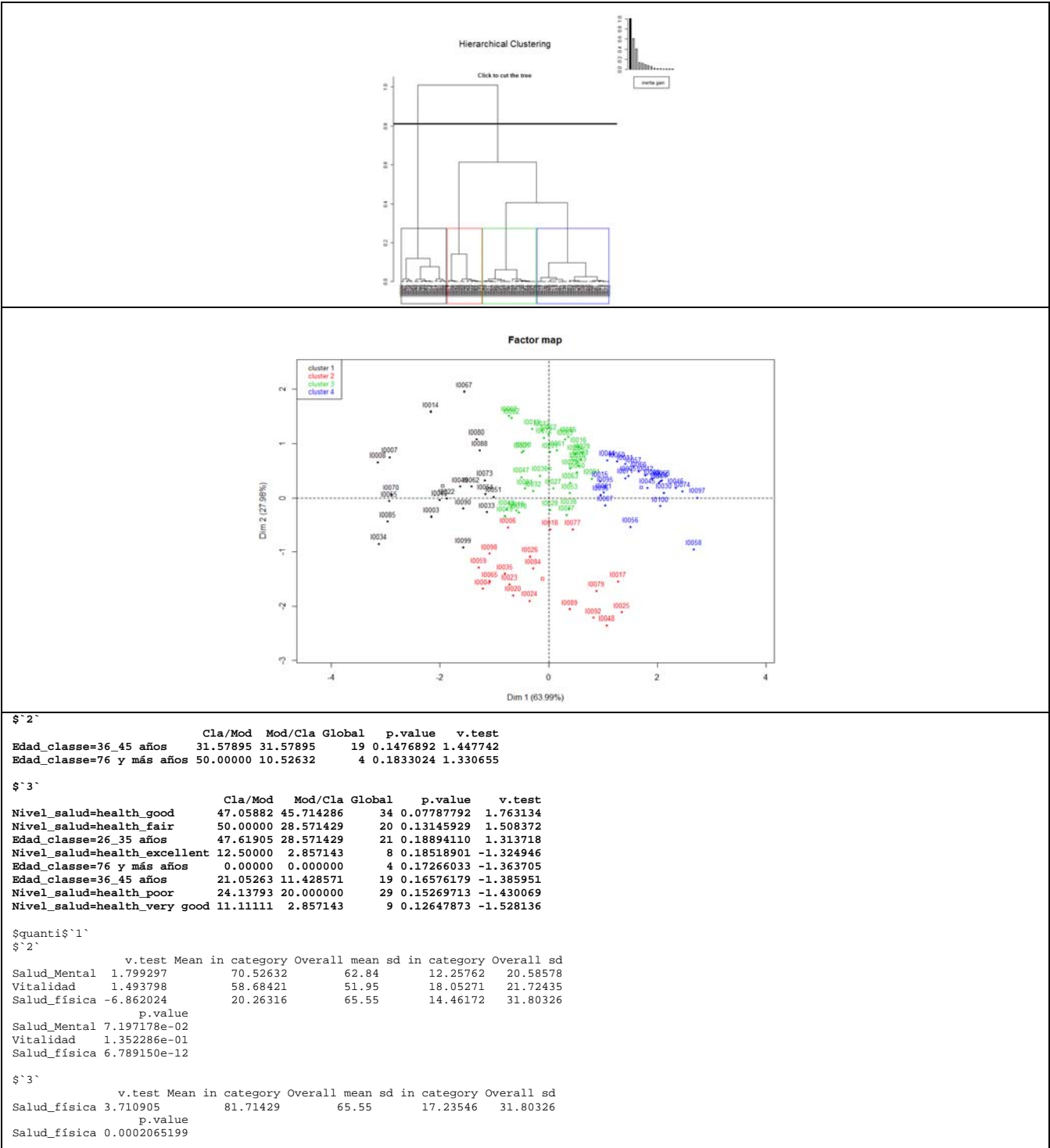


Figura 2. Principales resultados gráficos y numéricos de la clasificación automática