

# Comportamiento de la mortalidad en los países de la Unión Europea y caracterización según sus índices de mortalidad\*

**Luis Esteban Chaves**

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad  
Universitat Politècnica de València

**Ana María Debón Aucejo**

Centro de Gestión de la Calidad y del Cambio  
Universitat Politècnica de València

**M<sup>a</sup> Fulgencia Villa Juliá**

Grupo de Sistemas de Optimización Aplicada, Instituto Tecnológico de Informática  
Universitat Politècnica de València

## Resumen

---

Las comparaciones de longevidad en diferentes países ayudan a los responsables políticos y a los analistas a evaluar la cantidad de periodos de tiempo y de personas que se han beneficiado de los programas públicos, como la seguridad social y la asistencia sanitaria.

El objetivo fundamental de este trabajo es identificar grupos de países de la Unión Europea que tengan comportamientos similares en su mortalidad. Las diferencias entre los grupos serán cuantificadas a través de los indicadores de mortalidad.

El trabajo muestra que, teniendo en cuenta la mortalidad, los países de la Unión Europea se pueden agrupar en tres clusters. Uno formado por Alemania, Austria, Bélgica, Dinamarca, España, Finlandia, Francia, Holanda, Irlanda, Italia, Luxemburgo, Portugal, Reino Unido y Suecia. Otro formado por Bulgaria, Polonia, Eslovaquia, Eslovenia, República Checa y Hungría y un tercero integrado por Letonia, Estonia y Lituania. Esta agrupación es fruto del gap en mortalidad que existe entre los países del oeste de la Unión Europea y los del este, siendo en estos últimos donde se muestran valores más desfavorables. La caracterización de estos grupos muestra que la esperanza de vida al nacer es el indicador de mortalidad que más discrimina entre los clusters obtenidos.

---

\* Este trabajo ha sido financiado por un proyecto del Ministerio de Economía y Competitividad (MTM2013-45381-P).

*Palabras Clave:* Longevidad, Análisis de componentes principales, Análisis clúster, Árboles de decisión, Análisis de mortalidad, Unión Europea.

*Clasificación AMS:* 62H25, 62H30, 62G08

## **Patterns of mortality in European Union countries and caracterizacion according to their mortality indicators**

### **Abstract**

Comparisons of longevity by country help policymakers and analysts to evaluate the amount of individuals and time periods, in which they get benefits from public programs like social security and health care.

This research presents a method to group the European Union countries with similar mortality behavior. Mortality inequalities found in the groups of European Union countries are characterized by means of mortality indicators.

It is concluded that there are differences in mortality between East and West European Union, with a clear disadvantage to the East. So three clusters are found. The first one consists of Germany, Austria, Belgium, Denmark, Spain, Finland, France, Netherlands, Ireland, Italy, Luxembourg, Portugal, United Kingdom and Sweden. The second one consists of Bulgaria, Poland, Slovakia, Slovenia, Czech Republic and Hungary. Latvia, Estonia and Lithuania make up the last cluster. The characterization of these clusters shows that life expectancy at birth is the mortality indicator that discriminates the most among the clusters obtained.

*Keywords:* Longevity, Principal component analysis, Cluster analysis, Decision trees, Mortality analysis European Union.

*AMS Classification:* 62H25, 62H30, 62G08

## **1. Introducción**

La demografía es una ciencia que trata de cuantificar y analizar la dimensión, estructura, evolución y características generales de una determinada población. Disponer de información demográfica puede ser relevante para diferentes ámbitos, especialmente, económicos y sociales. En el estudio de O'Donnell (2009) se comenta que en Europa la desigualdad de la salud está ligada a las diferencias económicas entre clases sociales. Por lo tanto, el analizar la evolución de la desigualdad en el campo de la salud, permite evaluar y controlar si las políticas aplicadas tienen éxito para paliar esta desigualdad (Dalstra et al., 2002). Cuando se complementa con el campo de la medicina como lo hace Fries (1980), se puede entender por qué la esperanza de vida aumentó tanto durante el siglo XX. Básicamente, esto es debido a la disminución de las muertes prematuras por el avance en la medicina, al superar las muertes por causas infecciosas. Actualmente, todavía se aplican los modelos de envejecimiento y de enfermedad propuestos por el autor debido a su utilidad.

Uno de los aspectos más relevantes en el análisis demográfico es el estudio de la esperanza de vida y la mortalidad. Este es un tema que influye en diferentes ámbitos sociales ya que las modificaciones de la esperanza de vida implican tener que revisar políticas en materia de sanidad y seguridad social para que las mismas sean coherentes. Las instituciones gubernamentales, en base a esa clase de información, fijarán diferentes tipos de impuestos y políticas que garanticen la sostenibilidad en el tiempo del sistema de pensiones y de salud social (Zaidi, 2012). También, el estudiar la composición de una población, como lo hace la demografía, es útil para que los gobiernos intervengan en la planificación familiar y controlen el desarrollo generacional tanto a corto como a medio y largo plazo. Otro de los ámbitos, donde se analizan aspectos demográficos, es el actuarial, seguros y previsión. Las entidades aseguradoras utilizan la información para diseñar productos como: seguros de vida, seguros privados de salud, planes de pensiones, planes de jubilación... De tal manera, que sean atractivos para los usuarios y rentables para las compañías. Este diseño estará adaptado para cada país donde operen estas empresas.

En el último informe mundial de la OMS, publicado en septiembre de 2015 sobre el envejecimiento y salud, se apunta que la mayoría de las personas puede aspirar a vivir más allá de los 60 años y, en países con ingresos altos, donde se situaría la gran mayoría de países de la Unión Europea (UE), se aprecia un aumento sostenido de la esperanza de vida como consecuencia del descenso de la mortalidad en las personas mayores. Esta tendencia, combinada con las caídas en las tasas de fertilidad que se registran en casi todos los países, comienza a repercutir considerablemente en la estructura de las poblaciones. Ante esta perspectiva, los diferentes países que integran la UE deben ser conscientes de estos cambios y adoptar las medidas oportunas en diferentes materias como dar respuesta a temas relacionados con la salud pública o mantenimiento del estado de bienestar. Por tanto, uno de los grandes debates que se plantea, actualmente, es la sostenibilidad del Modelo Social Europeo (Bilbao, 2013) y uno de los puntos influyentes son los cambios demográficos que ha sufrido la Unión Europea tanto durante el siglo XX como lo que llevamos transcurrido del XXI. Según menciona Devesa et al. (2013), hay que reformar, en muchos casos, los obsoletos sistemas públicos de pensiones, debido, sobre todo, al envejecimiento de la población. Este envejecimiento hace peligrar la sostenibilidad del mismo a medio y largo plazo. Además, si no se aborda este aspecto demográfico se puede generar desconfianza en las generaciones cotizantes actuales al provocar incertidumbre sobre sus expectativas de jubilación. A pesar de la diversidad de países que integran la unión, algunos autores (Devesa et al., 2013) plantean la necesidad de afrontar los problemas demográficos de manera conjunta, en lugar de que cada país adopte sus reformas como ha venido sucediendo hasta ahora, con el objetivo de incrementar su efectividad.

Tal y como se ha mostrado, la aplicabilidad de la demografía y su importancia se extiende a muchos ámbitos. A partir de la demografía también se han desarrollado diferentes modelos, instrumentos e indicadores como el de vida potencial (Goerlich et al., 2013), las tablas de vida... que ayudan a entender, comparar y predecir la mortalidad para una toma de decisiones más objetiva en los diferentes campos donde se aplica.

Por todo lo expuesto anteriormente, consideramos interesante abordar un estudio de longevidad de todos los países de la Unión Europea y detectar diferentes características que puedan presentar sus integrantes. Dicho estudio puede ayudar a la hora de tomar decisiones para garantizar la sostenibilidad del Modelo Social Europeo. Varios estudios (Spinakis et al., 2011; Mackenbach et al., 1997; Mackenbach, 2013; Mackenbach et al., 2013; Dalstra et al., 2002; Meslé et al., 2002) han puesto de manifiesto las diferencias, en materia de salud, existentes entre países de la UE.

El objetivo fundamental de este trabajo es estudiar cómo se comporta la longevidad en los países de la UE, para identificar grupos de países que tengan comportamientos similares en cuanto a mortalidad. Para llevar a cabo esta identificación se aplicarán componentes principales y análisis de clúster jerárquico y fuzzy. Además, se pretende caracterizar, utilizando la técnica de árboles de decisión, estos grupos en base a los diferentes indicadores de mortalidad que mejor representen las desigualdades entre los países. Es vital tener en cuenta una característica como el género, ya que estudios como el de Leon (2011), Meslé et al. (2002), Meslé (2004) muestran que la mortalidad se comporta de forma diferente según analicemos hombres o mujeres. Los diferentes estudios muestran que la mortalidad en los hombres es más variable entre los países que para el caso de las mujeres. Así pues, nuestro estudio se centra en el análisis de la longevidad para el género masculino.

El presente documento está estructurado como sigue: en la Sección 2 se exponen las características y las fases que integran el estudio estadístico; en la Sección 3 se explica la metodología y los resultados obtenidos del análisis de componentes principales; en las Secciones 4 y 5 se detallarán la agrupación y caracterización de los países de la UE mediante la aplicación del análisis clúster y árboles de decisión, respectivamente. En la Sección 6 se mostrarán las conclusiones del estudio.

## **2. Características y fases del estudio estadístico**

El estudio analiza el fenómeno demográfico de la longevidad de los hombres en los países de la UE con la finalidad de identificar grupos de países que presentan comportamientos similares en cuanto a mortalidad. Además, muestra una caracterización de estas agrupaciones en base a indicadores demográficos relacionados con el fenómeno de la mortalidad.

A continuación, se detallan la fuente y las características de los datos que componen la información muestral y las fases que integran el trabajo.

La información para realizar el estudio ha sido obtenida de la página web The Human Mortality Database (HMD) ([www.mortality.org](http://www.mortality.org)). Esta página web fue creada, en el año 2000, para proporcionar datos detallados de mortalidad y poblaciones, principalmente para fines de investigación o de docencia (Human Mortality Database, 2015). Entre los estudios más recientes que han utilizado la información de la HMD se encuentran: estudios comparativos de la mortalidad como el de Booth et al. (2006), análisis de tendencia de la mortalidad como el de Fritzell et al. (2013) y el de Hatzopoulos et al. (2013).

Para realizar este estudio se ha utilizado la información que contienen las tablas de vida de periodo publicadas en la HMD. Estas tablas de vida son un instrumento de análisis demográfico, que pretenden describir el comportamiento del fenómeno de la mortalidad sobre una población durante un periodo definido de tiempo, y simula la incidencia del mismo sobre una cohorte o generación ficticia. Para el cálculo de las tablas de vida de periodo se parte de una cohorte ficticia inicial de 100.000 individuos a la que se le aplica un patrón de mortalidad definido. El cálculo de las funciones biométricas que muestra una tabla de vida se realiza a partir de las defunciones ocurridas en la población cada año.

Actualmente, la UE está integrada por 28 países. Sin embargo, en la HMD sólo están disponibles las tablas de vida de los siguientes 23 países:

Alemania	Eslovaquia	Holanda	Lituania	Reino Unido
Austria	España	Hungría	Luxemburgo	República Checa
Bélgica	Estonia	Irlanda	Polonia	Suecia
Bulgaria	Finlandia	Italia	Portugal	
Dinamarca	Francia	Letonia		

Los países que quedan fuera del estudio son Chipre, Croacia, Grecia, Malta y Rumania. No obstante, consideramos que la muestra formada por los 23 países mencionados es representativa del conjunto de los países que integran la UE, porque representan más del 80% de los integrantes. Además, los cinco países que quedan fuera son similares y están próximos geográficamente a otros que forman parte del estudio. Con lo cual la información que estos cuatro países podrían aportar está representada por otros similares a ellos.

Para el análisis de los datos muestrales se ha utilizado el software R Core Team (2015). R es un entorno de programación, análisis estadístico y de generación de gráficos distribuido gratuitamente. Además, la implicación de sus usuarios y continuas mejoras facilitan su uso. Actualmente, R está constituido por una gran cantidad de paquetes que permiten realizar diferentes tipos de análisis.

El estudio se estructura en dos fases fundamentales: agrupación de países según su longevidad y caracterización de grupos de países. A continuación, se expone la información muestral que se utiliza para desarrollar cada una de las fases.

#### a) Fase de agrupación

La agrupación de los 23 países se debe realizar atendiendo a alguna medida que cuantifique la mortalidad. Hay que señalar que, en general, entre los autores que realizan análisis de mortalidad, no hay una unanimidad a la hora de decidir la medida que mejor permite cuantificarla. Hay autores que para estudiar las diferencias en longevidad utilizan tasas de mortalidad como en Christiansen et al. (2015) y Fritzell et al. (2013). En este estudio se ha utilizado la probabilidad de muerte  $q_{xt}$  ya que ha sido una medida que ha dado buenos resultados en la modelización de la mortalidad (Debón et al. 2005, Currie 2016).

La probabilidad de muerte se define como la probabilidad, en un año determinado  $t$ , de que un individuo perteneciente a la cohorte inicial (grupo de individuos) con edad “ $x$ ” muera a su edad actual, es decir, que no llegue vivo a la edad “ $x + 1$ ”. El cálculo de esta función biométrica se muestra en la Ecuación 2-1. Señalar que  $l_{xt}$  indica, para el año  $t$ , la cantidad de supervivientes de la cohorte inicial, que llegan con vida a la edad “ $x$ ” y  $d_{xt}$  es el número de individuos, para el año  $t$ , que mueren en la edad “ $x$ ”, es decir no llegan a la edad “ $x + 1$ ”.

Ecuación 2-1

### Probabilidad de Muerte

$$q_{xt} = \frac{d_{xt}}{l_{xt}}$$

La probabilidad de muerte  $q_{xt}$  presenta problemas de heterocedasticidad por lo que hay que aplicarle una transformación. En este caso la más utilizada es la transformación logit (Ecuación 2-2)

Ecuación 2-2

### Transformación de $q_{xt}$

$$\log\left(\frac{q_{xt}}{1 - q_{xt}}\right)$$

la razón que lleva a utilizar esta transformación es que su inversa da como resultado para  $q_{xt}$  valores entre 0 y 1 respetando así, las cotas para la probabilidad.

Para esta primera fase se han utilizado las tablas de vida del periodo que abarca desde el 1990 al 2009 de los 23 países. La razón principal es que la mayoría de países tienen en este periodo la información completa en las tablas de vida. Hay países como Alemania que no disponen de información antes de 1990 y países como Eslovaquia, Eslovenia, Holanda, Hungría entre otros que sólo han actualizado la información hasta el año 2009. Como una de las ideas principales del estudio es identificar sus similitudes, necesitamos un periodo en el que la información sea la misma para todos, por esto se considera este periodo adecuado en cuanto a disponibilidad de información. Además, 20 años es un número de años muestral suficiente para analizar el comportamiento de la mortalidad en estos países. De hecho algunos autores como Meslé et al. (2002) en lugar de utilizar toda la información disponible durante un largo periodo de tiempo (entre los años 1965 y 1995 para los autores mencionados) sólo utilizan la información del primer y el último año. Señalar que las tablas de vida dan la información desglosada para cada año de vida, y estos años van de 0 a 110 años, por lo que tendremos el comportamiento de cada edad de una cohorte para cada año del periodo estudiado lo cual es bastante considerable. Señalar que la edad 110 engloba todos los individuos que tienen una edad igual a superior a 110 años.

Al realizar la transformación logit perdemos algunos datos ya que cuando  $q_{xt}$  toma los valores 0 o 1, por la composición de la fórmula utilizada, no se pueden calcular dichos valores, ya que no existe el logaritmo de 0, y tampoco es posible una división para 0 en caso de que  $q_{xt}$  tome el valor de 1. Por ejemplo, todos los datos correspondientes a la edad 110, por ser la última edad observada, tendrán una probabilidad de fallecimiento de 1, por lo tanto, todos los datos de la edad 110 se perderán y realmente sólo podremos trabajar con las edades de 0 a 109 años. Por tanto, tal y como muestra la Tabla 2-1, la muestra completa debería tener 50.600 datos pero debido a los datos faltantes hay 50.329 datos disponibles. Toda esta información se ha organizado en una matriz (Tabla 2-2) donde los individuos son los 23 países estudiados y las variables serán cada una de las probabilidades de muerte para cada edad observada hasta los 109 años de edad y para cada año del periodo de estudio definido que son 20 años.

Tabla 2-1

**Resumen de Datos de la Muestra**

	<i>Hombres</i>
Países	23
Edades (0 a 109)	110
Años (1990 a 2009)	20
Datos totales	50.600
Datos perdidos	271
<b>Datos disponibles</b>	<b>50.329</b>

Tabla 2-2

**Matriz de Datos Hombres**

<i>País</i>	1990-0	1990-1	...	1991-0	1991-1	...	2009-107	2009-108	209-109
Alemania	-4,8178	-7,3076	...	-4,9167	-7,3076	...	0,0722	0,1295	0,1828
Austria	-4,7580	-7,4348	...	-4,7592	-7,3534	...	0,0758	0,1335	0,1871
Bélgica	-4,6815	-7,2081	...	-4,6630	-7,6838	...	0,1269	0,1823	0,2335
Bulgaria	-4,0939	-6,4744	...	-3,9741	-6,2277	...	0,0107	0,0402	0,0885
Dinamarca	-4,7450	-7,1177	...	-4,8065	-7,2355	...	0,0583	0,1144	0,1667
.	.	.	...	.	.	...	.	.	.
.	.	.	...	.	.	...	.	.	.
.	.	.	...	.	.	...	.	.	.
.	.	.	...	.	.	...	.	.	.
Suecia	-4,9855	-7,8490	...	-5,0171	-7,9016	...	0,1700	0,2250	0,2754

Para realizar la agrupación en la primera fase se aplica el análisis clúster sobre las probabilidades de muerte para cada edad de 0 a 109 años y para un período de 20 años (1990-2010). Esto supone trabajar con 2200 variables para cada uno de los 23 países, lo que implica una cantidad muy elevada de variables. Por ello, con el objetivo de disminuir esta cantidad de variables con la menor pérdida de información posible, una fase previa a la identificación de grupos, es aplicar el análisis de componentes principales. En las secciones 3 y 4 se muestra con detalle el análisis estadístico aplicado.

#### b) Fase de caracterización

Una vez que se han identificado los grupos de países europeos en cuanto a la mortalidad se han caracterizado. Para alcanzar este objetivo, se ha aplicado la técnica de los árboles de decisión con indicadores demográficos relacionados con la mortalidad, que permiten definir el comportamiento en general de estos grupos. Se han seleccionado índices que permiten analizar tanto la tendencia como la variabilidad de la mortalidad y, además, suelen ser los más utilizados en diferentes publicaciones referentes a este tema (Meslé et al., 2002, Shkolnikov et al., 2003, Spinakis et al., 2011).

Esta caracterización es importante ya que permite detectar diferencias (Spinakis et al., 2011), y podría ayudar a la UE a afrontar el desafío de replantear su modelo social y hacerlo sostenible a medio y largo plazo (Bilbao, 2013).

Los índices que se han utilizado en la fase de caracterización son los siguientes:

- **Esperanza de vida:** representa el número de años que vivirá por término medio un grupo de individuos según la mortalidad que describe la tabla de vida (Leguina, 1973). Por tanto, es el número de años que por término medio le queda por vivir a un grupo de personas al nacimiento o también se puede calcular para una edad específica. La esperanza de vida resume un conjunto de tasas de mortalidad calculadas por edades o rangos de edades.

Ecuación 2-3

#### Esperanza de Vida para un Grupo de Personas

$$e_{xt} = \frac{T_{xt}}{l_{xt}},$$

donde  $T_{xt}$  es el número total de años que viven desde la edad  $x$  hasta la máxima edad alcanzable los supervivientes  $l_{xt}$  para el año  $t$ . Señalar que en este estudio hemos calculado la esperanza de vida al nacer y a los 65 años.

- **Edad modal de muerte:** Es una medida estadística que muestra la edad a la que muere el mayor número de personas en una población. Se selecciona esta medida porque tal y como señala Canudas-Romo (2008), esta es capaz de registrar cambios en la tendencia de la mortalidad que la esperanza de vida no puede.
- **Mortalidad infantil:** Es la probabilidad de muerte que tiene un recién nacido o de que este no llegue a cumplir un año de edad. Esta edad se considera crítica y por eso



la probabilidad es alta con respecto a los siguientes años de vida. Una vez cumplido el año de edad esta probabilidad desciende. Esta será la probabilidad de muerte  $q_{xt}$  cuando  $x = 0$  en un determinado año de defunción “ $t$ ”.

- **Coefficiente de Gini:** Este coeficiente está muy vinculado al ámbito económico para medir la desigualdad pero se puede adaptar al ámbito demográfico. En este caso medirá si la supervivencia está concentrada en un grupo de edades. Puede tomar valores entre 0 y 1, siendo 0 una perfecta distribución, es decir, todos los individuos morirían a la misma edad y, 1 una desigualdad total donde todos mueren a una edad muy temprana y una persona muere a una edad muy elevada. Uno de los enfoques más utilizados para su cálculo se muestra en la Ecuación 2-4 donde  $\omega$  es la última edad observada y  $l_0$  es la cohorte ficticia de 100.000 individuos.

Ecuación 2-4

### Coefficiente de Gini

$$G_t = \frac{\sum_{x=0}^{(\omega-1)} (f_{xt} - g_{xt})}{\sum_{x=0}^{(\omega-1)} f_{xt}}$$

$$\text{donde; } f_{xt} = 1 - \frac{l_{xt}}{l_{0t}}$$

$$\text{donde; } g_{xt} = \frac{T_{0t} - T_{xt} - xl_{xt}}{T_{0t}}$$

Donde  $f_{xt}$  es proporción de población acumulada y  $g_{xt}$  la proporción de años vividos por esa población.

Los indicadores señalados anteriormente se han calculado, partiendo de las tablas de vida de cada uno de los 23 países, para los años comprendidos entre 2005 y 2009. Se toma como primer año el 2005 ya que todos los países analizados, excepto Bulgaria, se encuentran dentro de la UE. Para profundizar en las propiedades de los indicadores mencionados y en su cálculo se recomienda leer Debón et al. (2012) y la bibliografía citada en él.

Por otra parte, buscamos que la caracterización diste lo menos posible de la fecha actual, por ello incluimos los datos hasta el 2009. Para llevar a cabo la caracterización de los grupos se aplica la técnica de árboles de decisión. En la sección 5 se muestra con detalle el análisis estadístico aplicado.

### 3. Análisis de Componentes Principales

Este estudio utiliza 2200 variables para cada país. Esta cantidad es muy elevada. Por ello, un paso previo a la identificación de grupos, es aplicar el análisis de componentes principales (PCA, Principal Component Analysis). El PCA a partir de  $n$  observaciones de  $p$  variables, analiza si es posible representar adecuadamente esta información con un

número menor de variables construidas como combinaciones lineales de las originales e identificadas como componentes principales y que además sean independientes. Hay diferentes técnicas para decidir la cantidad idónea de componentes con las que trabajar (Bro et al., 2014). Nosotros en este trabajo vamos a seleccionar el número de componentes que minimiza el error de validación cruzada. El lector interesado en la técnica PCA puede consultar el texto de Jackson (2005).

Señalar que esta técnica es sensible a la escala en la que se mida las variables. Pensemos que en ocasiones las variables vienen expresadas en diferentes medidas y es necesario realizar un proceso de estandarización de variables (Geladi et al., 1986). En nuestro estudio no aparece este problema porque todas las variables tienen la misma naturaleza, en este caso la transformación logit de la probabilidad de muerte para cada una de las edades en cada uno de los años.

Este análisis se ha realizado con el paquete de R *missMDA*, implementado por Husson y Josse (2014), el cual nos permite realizar un análisis de componentes principales y a su vez manejar datos faltantes, ya que como vimos en la sección 2 vamos a tener algunos datos perdidos debido a la transformación de la medida de mortalidad escogida.

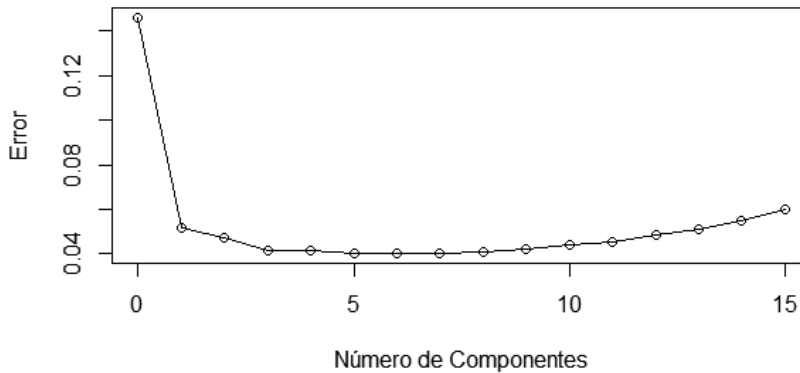
Para determinar la cantidad de variables latentes o componentes que utilizaremos para la fase de identificación de grupos se analiza la Tabla 3-1 y la Figura 3-1.

Tabla 3-1

### **Error Validación Cruzada Hombres**

<i>Componente</i>	<i>Error</i>
0	0,1461181
1	0,0518719
2	0,0471699
3	0,0415588
4	0,0416495
5	0,0404391
6	0,0404747
7	0,0407583
8	0,0415589
9	0,0425248
10	0,0445157
11	0,0462741
12	0,0496214
13	0,0526976
14	0,0568262
15	0,0620088

Figura 3-1

**Error Validación Cruzada Hombres**

Se observa en la Figura 3-1 que el error, según la técnica de validación cruzada, desciende más pronunciadamente hasta la componente 3, a partir de ahí, el error disminuye muy levemente hasta la componente 5 que es donde se observa el menor error y, a partir de ella, el error empieza a crecer. Por ello, el número de componentes seleccionado será 5, dado que corresponde al menor error (ver Tabla 3-1).

Una vez definido el número de componentes que vamos a utilizar para el modelo de PCA, vamos a analizar sus valores propios y la varianza explicada en la Tabla 3-2.

De acuerdo a la Tabla 3-2, necesitaríamos un total de 22 componentes para explicar el 100% de la varianza. Vemos que con la primera componente ya podemos explicar un 59% de la varianza total. Con las 5 componentes elegidas explicamos un 85% en total.

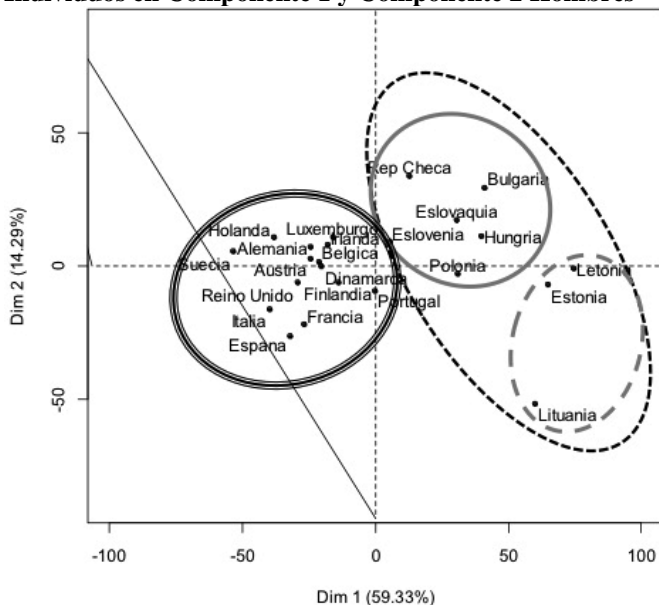
Para ver la situación de los países analizamos las 2 primeras componentes que son las que más varianza explican. En la Figura 3-2 se han representado sus scores o proyecciones.

Tabla 3-2

**Valores Propios y Varianza Explicada Hombres**

<i>Componente</i>	<i>Valor propio</i>	<i>Porcentaje de varianza</i>	<i>Porcentaje de Varianza Acumulado</i>
1	1.303,33	59,24	59,24
2	314,23	14,28	73,53
3	146,57	6,66	80,19
4	70,29	3,20	83,38
5	53,69	2,44	85,82
6	48,08	2,19	88,01
7	37,64	1,71	89,72
8	31,98	1,45	91,17
9	30,82	1,39	92,56
10	23,83	1,08	93,64
11	22,26	1,01	94,66
12	19,32	0,88	95,53
13	17,24	0,78	96,32
14	16,64	0,76	97,07
15	14,03	0,64	97,71
16	12,82	0,58	98,29
17	10,62	0,48	98,78
18	7,90	0,36	99,14
19	6,23	0,28	99,42
20	4,96	0,23	99,64
21	4,29	0,19	99,84
22	3,53	0,16	100,00

Figura 3-2

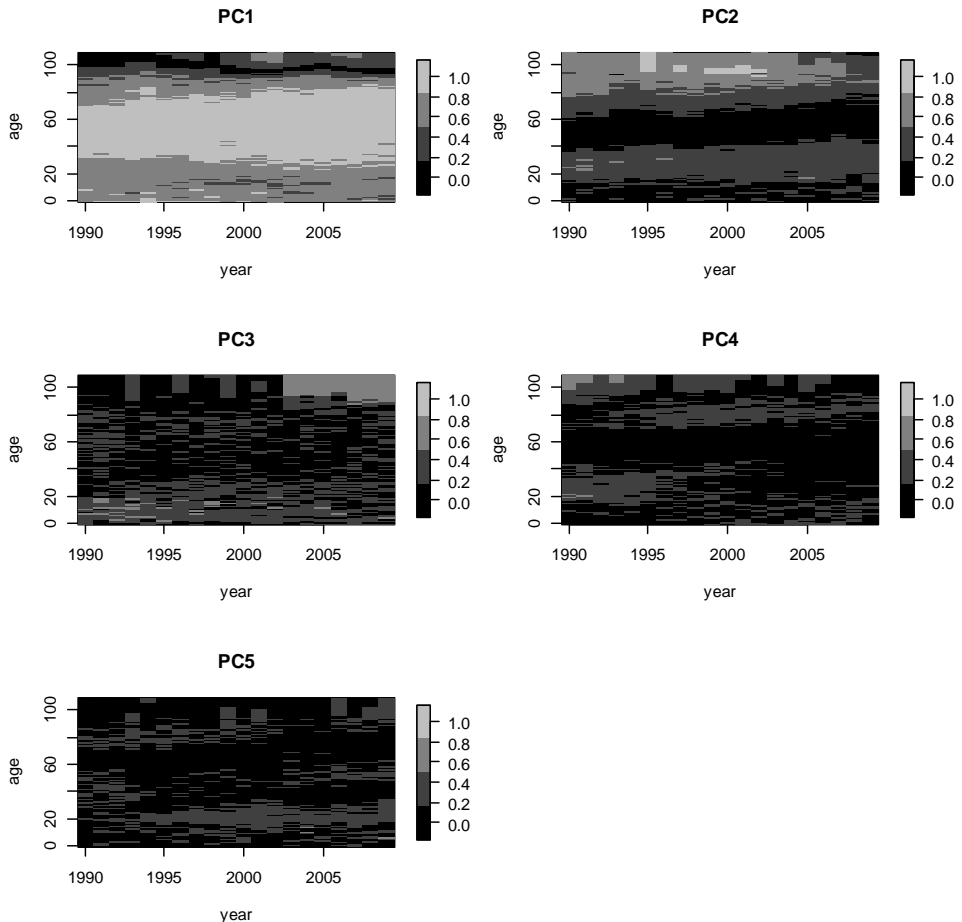
**Individuos en Componente 1 y Componente 2 Hombres**

Se observan 2 grandes grupos. Uno enmarcado con un círculo triplemente remarcado que engloba a países del Oeste de Europa y otro enmarcado con un círculo negro discontinuo integrado por países del Este de Europa. El grupo de los países del Oeste están más próximos entre sí que los países del Este. Esto significa que tienen un comportamiento más parecido entre ellos. Sin embargo, los países del Este se pueden dividir a su vez en otros 2 grupos, siendo Letonia, Estonia y Lituania los países que presentan un comportamiento diferente al resto de países del Este. Este grupo de países bálticos presentan una mortalidad superior por lo que se encuentran más a la derecha del eje x.

Hemos alcanzado el objetivo de reducir la dimensión de las 2200 variables que teníamos, tratando de perder lo mínimo de información posible. Trabajaremos con 5 componentes las cuales explican aproximadamente el 86% de la varianza. De las variables latentes definidas, usaremos sus scores, que es la información de nuestros 23 individuos en cada componente.

Para conseguir una interpretación de las componentes, la Figura 3-3 muestra los coeficientes que relacionan las componentes principales con cada una de las probabilidades de muerte de cada edad y año. Como puede observarse la primera componente está relacionada con las edades de 0-85 para todos los años, la componente 2 está relacionada con las edades elevadas de 80 hasta 109 para los primeros años de 1991 hasta 2005-2007, mientras que la componentes 3 completa tiene relación con la edades elevadas de 85 a 109 para los últimos años de 2007 a 2009. Hasta esta tercera componente la variabilidad explicada era aproximadamente del 80%, el resto de componentes explican solo el 8% y no tienen una interpretación clara.

Figura 3-3

**Relación de las variables con las componentes principales**

El siguiente paso es utilizar esta información para identificar países que tengan comportamientos similares. Hemos visto ya a priori que se han ido formando grupos en la Figura 3-2, pero realmente hay que validar estas agrupaciones. Para esto aplicamos el análisis clúster tal como se muestra en la siguiente sección. Gracias a ello, se identificará como de compactos son estos grupos según sus distancias y si realmente estas son las agrupaciones más adecuadas.

**4. Análisis Clúster**

La fase de identificación de grupos se ha realizado mediante el análisis clúster. Este análisis clasifica los individuos en grupos relativamente homogéneos en base a las

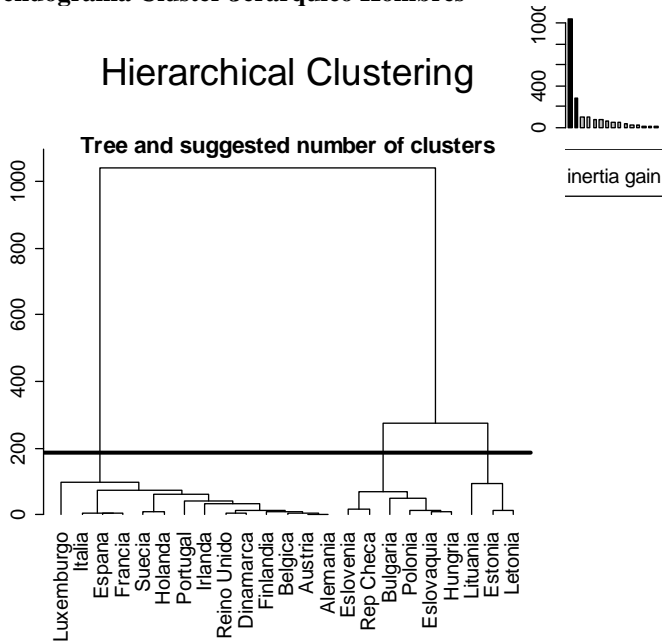
variables analizadas. Dentro de cada grupo o clúster, se debe tener individuos similares entre sí, es decir, se debe tener una alta homogeneidad interna, además ser diferentes a los de otros grupos, es decir, alta heterogeneidad externa. Hay diferentes métodos de análisis clúster pero en este estudio se han utilizado dos con el objetivo de comparar la coherencia de sus agrupaciones. De esta forma se verifica la robustez de los resultados. Los dos métodos utilizados han sido:

- **Clúster Jerárquico:** es uno de los métodos más usados ya que sus resultados son bastante estables. Se aplica el método de Ward como algoritmo de resolución, ya que comparado con otros es el que genera los mejores resultados en la mayoría de casos (Leschke, 2005). Este fue diseñado para minimizar la varianza dentro de los grupos usando la suma de errores al cuadrado. Este método suele conseguir grupos lo más homogéneos posibles. Además, se utiliza el dendograma, la herramienta que ayuda a visualizar las particiones y las distancias entre los grupos, por lo cuál es de gran ayuda a la hora de decidir el número más idóneo de clústers (Leschke, 2005).
- **Fuzzy clúster:** En los análisis clásicos de clúster cada individuo pertenecerá a un solo clúster. El algoritmo fuzzy clustering se caracteriza por el hecho de que los individuos pueden pertenecer a más de un grupo. Esa asociación a cada grupo está medida por un nivel de pertenencia. Este nivel de pertenencia indicará la fuerza de relación entre el individuo y un grupo en particular. El algoritmo más usado para esta técnica es el fuzzy C-Means (Pal et al., 1996). Esta técnica fue originalmente introducida por Jim Bezdek como una mejora sobre los métodos de clustering (Bezdek et al., 1981).

Para el análisis clúster jerárquico se ha utilizado el paquete de R **FactorMineR**, implementado por Husson et al. (2015) y el paquete **e1071**, diseñado por Meyer et al. (2014), para el fuzzy clúster. Ambos análisis clúster se han realizado con las cinco variables latentes resultantes del PCA.

La figura 4-1 muestra el dendograma obtenido tras aplicar el clúster jerárquico. Esta representación sugiere una agrupación de países en 3 grupos. Dicha agrupación está basada en la distancia euclídea entre los países.

Figura 4-1

**Dendrograma Clúster Jerárquico Hombres**

Analizando de forma detallada el dendrograma (Figura 4-1), se observa que podemos agrupar los países en 2 o 3 clústers. En la Tabla 4-1 se muestran los países que integrarían la división en 2 o en 3 clústers. Comparando estas agrupaciones con lo que se había obtenido en el análisis de componentes principales (Figura 3-2), observamos que se obtendría las mismas agrupaciones-



Tabla 4-1

**Integrantes con 2 Clústers y 3 Clústers**

<i>2 Cluster</i>		<i>3 Cluster</i>		
<i>Grupo 1</i>	<i>Grupo 2</i>	<i>Grupo 1</i>	<i>Grupo 2</i>	<i>Grupo 3</i>
Alemania	Bulgaria	Alemania	Bulgaria	Estonia
Austria	Eslovaquia	Austria	Eslovaquia	Letonia
Bélgica	Eslovenia	Bélgica	Eslovenia	Lituania
Dinamarca	Estonia	Dinamarca	Hungría	
España	Hungría	España	Polonia	
Finlandia	Letonia	Finlandia	Rep. Checa	
Francia	Lituania	Francia		
Holanda	Polonia	Holanda		
Irlanda	Rep. Checa	Irlanda		
Italia		Italia		
Luxemburgo		Luxemburgo		
Portugal		Portugal		
Reino Unido		Reino Unido		
Suecia		Suecia		

Para entender estas particiones es interesante analizar la historia de la UE en cuanto al proceso de incorporación de nuevos miembros desde su fundación en 1957. Revisando la información de la página web oficial de la UE (europa.eu) podemos ver que la cronología de la incorporación de los países ha sido la siguiente:

- Fundadores (1957): Alemania, Bélgica, Francia, Italia, Luxemburgo y Países Bajos.
- Primera ampliación (1973): Dinamarca, Irlanda y Reino Unido.
- Segunda ampliación (1981): Grecia.
- Tercera ampliación (1986): España y Portugal.
- Cuarta ampliación (1990): Reunificación alemana.
- Quinta ampliación (1995): Austria, Finlandia y Suecia.
- Sexta ampliación (2004): Chipre, Eslovaquia, Eslovenia, Estonia, Hungría, Letonia, Lituania, Malta, Polonia y República Checa.
- Séptima ampliación (2007): Bulgaria y Rumanía.
- Octava ampliación (2013): Croacia.

El periodo que se analiza en el trabajo es de 1990-2009, en el que hay una importante ampliación de la UE, pues es en el año 2004 donde se unen países que en su mayoría vienen de un régimen económico muy distinto al imperante entre los países de la UE. Estos países, que son los de Europa del Este, como lo explica Leon (2011), por diferentes razones políticas, económicas y hábitos de su vida cotidiana, presentan una

menor longevidad. A partir de la disolución de la Unión Soviética estos países han pasado por un periodo de transición hasta unirse a la UE en 2004, donde han tenido que modificar diferentes criterios para ser aceptados por los países miembros. Este proceso ha tenido un impacto positivo sobre la salud. Sin embargo, a pesar de que estos países han mejorado en este ámbito, todavía hay una brecha importante entre los países del Este y el Oeste.

Volviendo a las agrupaciones identificadas en la Figura 4-1, al pasar de 2 clústers a 3 clústers lo único que se genera es una partición en el segundo grupo, ya que hay 3 países que son Estonia, Letonia y Lituania que tienen un comportamiento más similar entre ellos. A continuación, con el fuzzy clúster contrastamos las situaciones de 2 y 3 grupos, para ello se calculan las probabilidades de pertenencia de los países a los diferentes grupos. Esta información se muestra en la Tabla 4-2 para 2 y 3 clústers, respectivamente.

Tabla 4-2

**Probabilidad de Pertenencia Fuzzy Clúster con 2 y 3 Grupos Hombres**

<i>Cluster</i>	<i>1</i>	<i>2</i>	<i>Cluster</i>	<i>1</i>	<i>2</i>	<i>3</i>
Alemania	96,92%	3,08%	Alemania	92,35%	5,91%	1,74%
Austria	97,20%	2,80%	Austria	93,25%	5,08%	1,67%
Bélgica	98,66%	1,34%	Bélgica	94,97%	3,80%	1,22%
Bulgaria	12,83%	87,17%	Bulgaria	8,58%	70,24%	21,18%
Dinamarca	95,96%	4,04%	Dinamarca	90,09%	7,65%	2,26%
Eslovaquia	12,08%	87,92%	Eslovaquia	3,30%	91,52%	5,18%
Eslovenia	58,24%	41,76%	Eslovenia	25,82%	65,94%	8,24%
España	90,24%	9,76%	España	83,71%	10,81%	5,48%
Estonia	10,57%	89,43%	Estonia	1,46%	4,88%	93,66%
Finlandia	95,63%	4,37%	Finlandia	87,64%	9,42%	2,94%
Francia	90,14%	9,86%	Francia	82,86%	11,55%	5,59%
Holanda	92,25%	7,75%	Holanda	85,21%	10,88%	3,91%
Hungría	12,14%	87,86%	Hungría	9,52%	68,57%	21,91%
Irlanda	82,93%	17,07%	Irlanda	66,60%	25,01%	8,38%
Italia	94,06%	5,94%	Italia	90,67%	6,38%	2,95%
Letonia	12,45%	87,55%	Letonia	2,55%	8,12%	89,32%
Lituania	27,42%	72,58%	Lituania	13,65%	23,32%	63,03%
Luxemburgo	71,27%	28,73%	Luxemburgo	53,26%	31,47%	15,27%
Polonia	7,49%	92,51%	Polonia	7,88%	76,72%	17,40%
Portugal	72,94%	27,06%	Portugal	53,61%	33,27%	13,12%
Reino Unido	97,11%	2,89%	Reino Unido	95,00%	3,65%	1,35%
Rep. Checa	40,68%	59,32%	Rep. Checa	15,37%	74,97%	9,66%
Suecia	90,68%	9,32%	Suecia	84,03%	11,13%	4,84%

Para el análisis clúster de 2 grupos, las probabilidades más altas de pertenencia de los países están en los mismos grupos que habíamos definido según el clúster jerárquico con excepción de Eslovenia que pasaría al primer grupo, teniendo una probabilidad de pertenencia de 58% para el primer grupo y de 42% para el segundo, señalar que su porcentaje de pertenencia es el más bajo de todos los países de su grupo.

Por otro lado, en el análisis clúster de 3 grupos, las probabilidades más altas de pertenencia de los países están exactamente en los mismos grupos que habíamos definido según el clúster jerárquico sin excepción. Aquí Eslovenia se sitúa claramente en el segundo grupo con un 65% de probabilidad de pertenencia.

Con el objetivo de decidir con qué partición quedarnos se analiza, una serie de índices (Rezaee et al., 1998) que ayudan a identificar lo compactos que son los grupos. Estos valores se muestran en la Tabla 4-3.

Tabla 4-3

#### **Índices de Validación de Clústers Hombres**

<i>Cluster</i>	2	3
xb	0,01	0,01
fs	-19608	-27036
pc	0,79	0,69
pe	0,35	0,56

Los índices xie beni (*xb*), fukuyama sugeno (*fs*) y partition entropy (*pe*) muestran que la agrupación es mejor cuanto más bajos sean sus valores. En *xb* las 2 agrupaciones son igual de buenas, *fs* muestra mejores resultados para 3 clústers y en *pe* una mejor agrupación para 2 clústers. En el índice partition coefficient (*pc*) valores mayores mostrarán una mejor partición, por lo que la mejor agrupación es la de 2 clústers. En general, aunque las 2 agrupaciones son bastante buenas, decidimos seguir trabajando con 3 grupos porque es la más coherente con los resultados obtenidos por componentes principales.

A continuación, localizaremos los países en el mapa de la UE con la finalidad de obtener conclusiones a la vista de su ubicación geográfica.

Figura 4-2

**Mapa de la Unión Europea para Clúster de Hombres**

Según la Figura 4-2 todos los países pertenecientes a la Unión Europea son los países de color gris; la línea de color azul es la división entre los 2 primeros grupos del clúster. Por tanto, los países del Oeste de Europa tienen un comportamiento diferente a los países del Este, como señalan trabajos sobre la esperanza de vida en Europa (Leon, 2011). Dentro del grupo del Este observamos otra partición que son los países del círculo rojo, que son 3 países que tienen un comportamiento de mayor mortalidad que el resto de Europa del Este, estos son conocidos como países bálticos por su ubicación (Leon, 2011). También podemos ver que Eslovenia está situada debajo de Austria y al

lado de Italia, muy cercano al Oeste, lo que hace que su pertenencia a un solo grupo sea más ambigua que otros países.

## 5. Árboles de decisión

Los árboles de decisión son una técnica tanto de clasificación como de regresión muy utilizada. Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, así para llegar a tomar la decisión final se sigue un conjunto de condiciones organizadas que se cumplen desde la raíz del árbol hasta sus diferentes ramas. Una de sus ventajas, es que debido a su forma, las posibles opciones a partir de una determinada condición son excluyentes. Esto permite analizar una situación y llegar a una sola acción o decisión a tomar. La tarea de aprendizaje para la cual los árboles tienen mejores resultados es para la clasificación (Orallo et al., 2004). Para el aprendizaje del árbol se usan algoritmos de partición y uno de los más conocidos es el CART (Breiman et al., 1984).

Para aplicar esta técnica hay que determinar el número de particiones que va a tener el árbol y el criterio de selección de las particiones (poda del árbol). Las particiones son un conjunto de condiciones exhaustivas y excluyentes. Cuantas más condiciones se permitan, más probabilidades existen de encontrar los patrones que hay en los datos y, en consecuencia, serán más precisos. Hay dos tipos de particiones, nominal y numérica. En la partición nominal el atributo tiene un número finito de valores que están ligados a una característica en los datos. En la numérica el atributo también es numérico pero continuo. Tras construir el árbol, éste se somete a un proceso de poda cuya finalidad es simplificar el árbol. Gracias a la poda se evita que el modelo esté sobreajustado, lo que perjudicaría el comportamiento global del modelo.

Las características numéricas que se utilizan para crear el árbol y así poder caracterizar los grupos identificados en el análisis clúster son: la esperanza de vida, esperanza de vida a los 65 años, moda, mortalidad infantil y el coeficiente de Gini. Todos estos indicadores se calculan para el periodo comprendido que va desde el 2005 hasta el 2009. Por tanto, al tener los grupos caracterizados podremos compararlos entre si y saber cómo se encuentran en cuanto a mortalidad de forma sencilla, basándonos en los indicadores anteriores.

La Tabla 5-2 es un ejemplo de cómo se ha organizado la información. Los indicadores se calcularon para todos los años definidos. Adicionalmente se añade una segunda columna con el clúster al que pertenece cada país.

Tabla 5-2

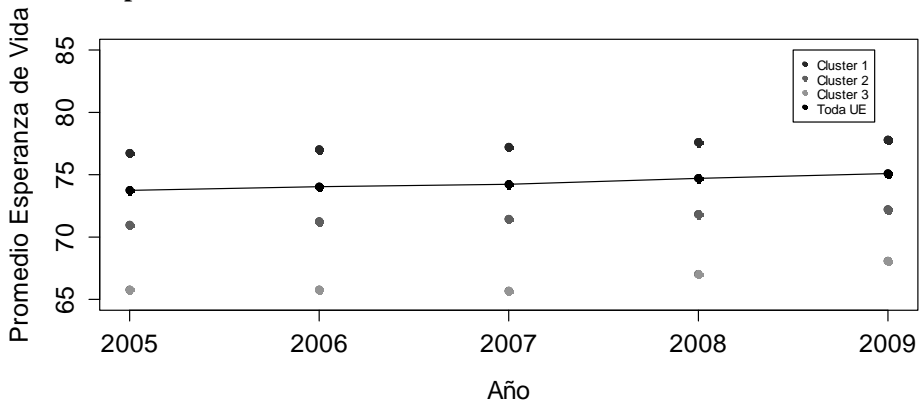
**Indicadores de Mortalidad de Alemania y Austria para 2009 en Hombres**

<i>País</i>	<i>Cluster</i>	<i>Gini-09</i>	<i>Esperanza de vida - 09</i>	<i>Esperanza de vida 65 años - 09</i>	<i>Moda - 09</i>	<i>Mortalidad Infantil - 09</i>
Alemania	1	0,0820	77,4600	17,2800	85,0000	0,0039
Austria	1	0,0865	77,4300	17,4900	86,0000	0,0040

Antes de aplicar la metodología de los árboles de decisión se ha realizado un análisis descriptivo del comportamiento de los indicadores durante los cinco años analizados. Este análisis ha sido aplicado para el conjunto de los 23 países que integran la muestra y para los clústers definidos. Del análisis realizado podemos destacar que para los hombres la esperanza de vida es el indicador que más ha mejorado. Esta ha incrementado en promedio 2 años del 2005 al 2009, de igual manera se observa el incremento sostenido conforme pasan los años. La esperanza de vida a los 65 años y la moda también han mejorado pero con menor intensidad que la esperanza de vida al nacer. La mortalidad infantil y el índice de Gini han disminuido ligeramente.

La Figura 5-1 muestra el promedio de la esperanza de vida del conjunto de países y de cada clúster definido. Se observa que el primer clúster está por encima del promedio de los 23 países, mientras que los otros 2 clústers están por debajo. También es importante señalar que el tercer clúster formado por los países bálticos, desde su ingreso en la UE ha incrementado su esperanza de vida, disminuyendo la brecha que tenía al principio con los otros 2 grupos.

Figura 5-1

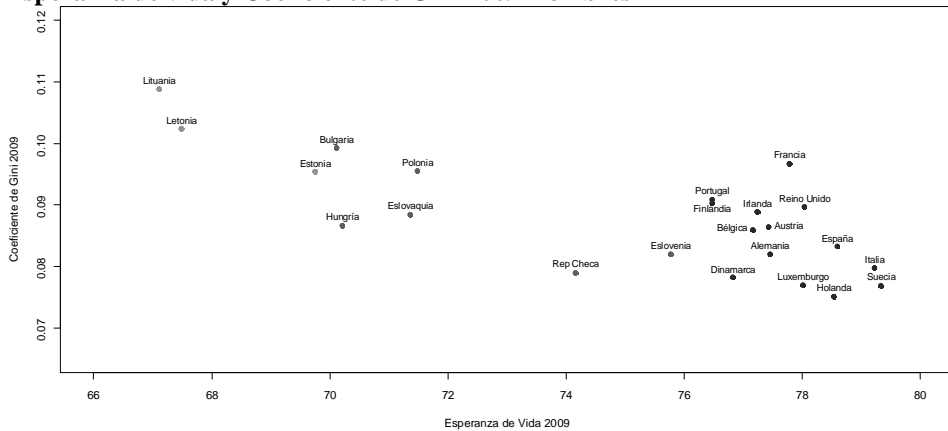
**Promedio por Clúster Hombres 2005 – 2009**

Si analizamos la Figura 5-2 que relaciona la esperanza de vida y el coeficiente de Gini para el último año de los 23 países analizados, podemos ver que la esperanza de vida

forma 3 grupos, que coinciden con los clústers definidos para los hombres anteriormente. Lituania, Letonia y Estonia tienen una esperanza de vida menor a 70 años. Bulgaria, Hungría, Polonia, Eslovaquia, República Checa y Eslovenia tendrán una esperanza de vida mayor a 70 años y menor a 76 años. El último grupo tendrá una esperanza de vida mayor a 76 años. Mientras que el Coeficiente de Gini no muestra discriminación entre los grupos. Sin embargo, cabe destacar que los países que tienen mayor Coeficiente de Gini que son Lituania y Letonia también muestran una menor esperanza de vida.

Figura 5-2

**Esperanza de vida y Coeficiente de Gini 2009 Hombres**



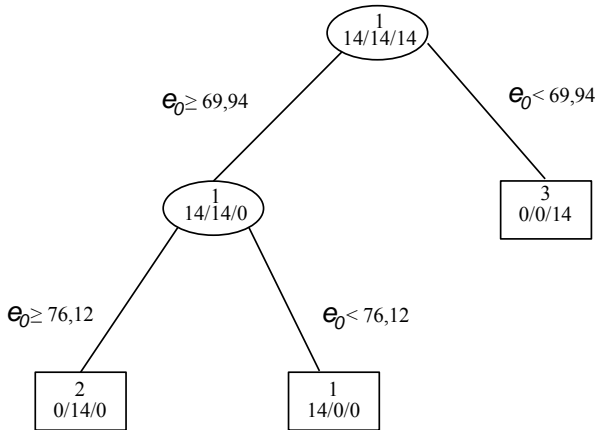
El siguiente paso es caracterizar los clústers definidos usando los diferentes indicadores de mortalidad mediante los árboles de decisión. La técnica de árboles de decisión está disponible en el paquete de **R rpart**, que ha sido implementado por Therneau, Atkinson y Ripley (2015).

Para aplicar esta técnica es necesario que las muestras estén equilibradas. Con el objetivo de disminuir el error, esta técnica tiende a ubicar a los individuos en el grupo más numeroso, lo que puede provocar que se pierdan clasificaciones. Por tanto, lo primero que haremos será equilibrar la muestra de los tres clústers de hombres formados por 14, 6 y 3 países, equilibrando la muestra correspondiente a los grupos 2 y 3. A estos les aplicaremos el muestreo con reemplazamiento para conseguir grupos formados por el mismo número de países. De esta manera conseguimos tres grupos con 14 países cada uno con lo que tenemos una base de datos formada por 42 países. Señalar que el muestreo con reemplazamiento consiste en que a partir de una muestra inicial se crean individuos adicionales basados en la información ya disponible. De esta manera se repetirán las observaciones originales aleatoriamente las veces que sean necesarias. Es una técnica adecuada para realizar determinados análisis, sobre todo discriminantes, donde se necesita tener muestras equilibradas. Si bien esto da buenos resultados hay que

tener en cuenta que no es bueno abusar de esta técnica cuando la varianza en los datos es elevada, estos análisis se realizarán con el software **R Core Team** (2015).

Figura 5-3

### Árbol de Clasificación de Hombres con Todos los Indicadores



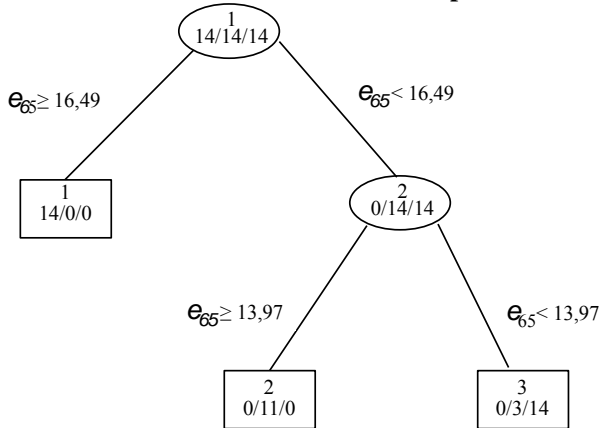
Podemos ver en la Figura 5-3 que la clasificación sólo utiliza la esperanza de vida al nacer y agrupa los países correctamente. Si observamos el árbol, este tiene 2 nodos, el ovalado, conocido como nodo de incertidumbre, aquí se siguen realizando particiones y el rectangular, conocido como nodo de decisión, donde el árbol arroja el resultado final de una ramificación. En el nodo de decisión del clúster 1, se muestra que el árbol agrupó correctamente los 14 individuos que le correspondían y no confundió ningún individuo del clúster 2, ni del clúster 3. Lo mismo que se observa en el clúster 1 sucede con el segundo y el tercer clúster.

Los miembros del grupo 3 tendrán una esperanza de vida menor a 69,94 años, los del grupo 2 mayor o igual a 69,94 años y menor a 76,12 años y los del primer grupo tendrán una esperanza de vida mayor o igual a 76,12 años. Debido a que la esperanza de vida es un indicador con un fuerte poder de clasificación en esta técnica, no permite que los otros indicadores aparezcan, ya que no son necesarios para la clasificación. Por lo que procederemos a realizar otro árbol de clasificación sin este indicador (Figura 5-4) para poder ver más características en los grupos.



Figura 5-4

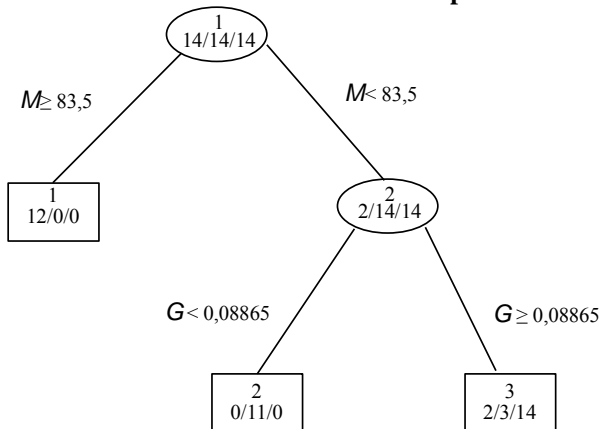
**Árbol de Clasificación de Hombres sin Esperanza de Vida**



Quitando la esperanza de vida al nacer, la agrupación ya no es tan correcta (Figura 5-4), pues agrupa 3 países del clúster 2 en el clúster 3. Sólo en el clúster 1 no hay confusiones. La interpretación de los resultados del árbol de la Figura 5-4 la podemos realizar de igual manera que hicimos con el primer árbol, los miembros del grupo 1 tendrán una esperanza de vida a los 65 años igual o superior a 16,49 años, los miembros del grupo 2 mayor o igual a 13,97 años y menor a 16,49 años y los del tercer grupo que son los que tendrán una esperanza de vida a los 65 años menor a 13,97 años. De igual manera volveremos a realizar otro árbol excluyendo este indicador para ver qué sucede, este se muestra en la Figura 5-5.

Figura 5-5

**Árbol de Clasificación de Hombres sin Esperanza de Vida Total ni a los 65 Años**



Quitando los 2 indicadores de esperanza de vida, la agrupación tiene una menor precisión, ya que agrupa 3 países del clúster 2 en el clúster 3 y 2 países del clúster 1 en el clúster 3. En este árbol de la Figura 5-5, los miembros del grupo 1 tendrán una moda igual o superior a 83,5 años, los miembros del grupo 2 y 3 menor a 83,5 años. La clasificación entre el clúster 2 y 3 se hace mediante el coeficiente de Gini, los que tengan un valor inferior al 0,08865 pertenecerán al clúster 2 y los que presenten un valor mayor o igual de 0,08865 al clúster 3. Señalar que la mortalidad infantil no ayuda a discriminar entre grupos.

Para resumir la caracterización de los grupos con todo lo revisado anteriormente mostramos la Tabla 5-3.

Tabla 5-3

### Caracterización Hombres

<i>Caracterización Hombres</i>				
<i>Indicador</i>	<i>Grupo 1</i>	<i>Grupo 2</i>	<i>Grupo 3</i>	<i>Error</i>
Esperanza	Mayor 76,12	Entre 76,12 y 69,94	Menor 69,94	0%
Esperanza 65	Mayor 16,49	Entre 16,49 y 13,97	Menor 13,97	7%
Moda	Mayor 83,50	Menor 83,50	Menor 83,50	12%
Gini		Menor 0,088	Mayor 0,088	

Los mejores indicadores para caracterizar los grupos son la esperanza de vida al nacer y la esperanza de vida a los 65 años, siendo mejor la esperanza de vida al nacer. Tras caracterizar los grupos se ha observado (Tabla 5-2) que los países de la UE pertenecientes a la parte Este y los que pertenecen a la zona Oeste presentan comportamientos diferentes. Además, el grupo de los países Bálticos formado por Estonia, Letonia y Lituania tienen un comportamiento diferente al resto de países del Este y, además, es el que presenta los peores valores para los indicadores de mortalidad analizados.

## 6. Conclusiones

En este trabajo se ha analizado la longevidad de los hombres, encontrado grupos de países de la Unión Europea que tienen un comportamiento parecido en cuanto a mortalidad y se ha logrado caracterizarlos mediante indicadores de mortalidad.

Primero, se ha recopilado información de mortalidad de 23 países de la UE de 0 a 110 años de edad, para el periodo 1990 – 2009. Seguidamente, para reducir la dimensión de esta información se ha aplicado el análisis de componentes principales. Esta técnica nos ha permitido comprimir la información de los hombres en 5 componentes, explicando aproximadamente el 86% de la información total.

Una vez comprimida la información en un número de variables más manejable, mediante el análisis clúster, se ha conseguido identificar grupos de países con

comportamientos similares en mortalidad. Se han dividido los países en los siguientes tres 3 grupos:

- Uno en el Oeste de Europa formado por: Alemania, Austria, Bélgica, Dinamarca, España, Finlandia, Francia, Holanda, Irlanda, Italia, Luxemburgo, Portugal, Reino Unido y Suecia.
- Dos en el Este. Uno formado por los tres países bálticos que formaban parte de la antigua URSS: Letonia, Estonia y Lituania. Y otro compuesto por: Bulgaria, Polonia, Eslovaquia, Eslovenia, República Checa y Hungría.

Antes de pasar a la caracterización, una vez definidos los grupos, se ha calculado para los años 2005-2009 los siguientes indicadores de mortalidad: esperanza de vida al nacer, la esperanza de vida a los 65 años, el índice de Gini, la moda y la mortalidad infantil. Con toda la información proporcionada por los citados indicadores se ha realizado un análisis descriptivo para los 23 países que integran nuestro estudio y también para los clústers. Este análisis nos ha permitido identificar tendencias y diferencias que existen entre los grupos. Se ha observado que la mortalidad ha mejorado con el paso del tiempo ya que todos los indicadores tanto de tendencia como de dispersión han mejorado. La esperanza de vida al nacer ha sido el indicador que más ha mejorado durante el periodo 2005-2009. Esta se ha incrementado 2 años por término medio. La esperanza de vida a los 65 años, la moda, la mortalidad infantil y el índice de Gini también han mejorado pero en menor intensidad que la esperanza de vida al nacer.

Por último, se han aplicado árboles de decisión para caracterizar los clústers identificados. Con esta técnica se ha logrado describir el comportamiento de los grupos identificados según los indicadores de mortalidad del año 2009. Gracias a la caracterización proporcionada por esta técnica se ha observado que:

- La longevidad de los hombres muestra diferencias entre los países del Este y Oeste de Europa, pudiendo observar una mayor esperanza de vida en los países del Oeste. La población con mayor longevidad analizada, serán los hombres de Europa del Oeste, con una esperanza de vida superior a los 76 años.
- En los hombres, dentro de los países que conforman el clúster del Este, hay un grupo de 3 países que son Estonia, Letonia y Lituania, que tienen una longevidad menor a la del resto de países de su grupo. Por lo tanto, también muestran la menor longevidad entre los países que conforman la Unión Europea.
- Hay que destacar, que los indicadores de mortalidad que mejor caracterizan a estos grupos de países son: la esperanza de vida al nacer, a los 65 años edad y la moda.
- La mortalidad infantil y el coeficiente de Gini aportan información que discrimina países de forma menos exhaustiva que los indicadores esperanza.
- Por último, el coeficiente de Gini toma valores en general bajos, lo que indica en general que la distribución de la mortalidad es equitativa en toda la Unión Europea.

Finalmente cabe señalar limitaciones de este estudio pues en base a todo lo analizado y concluido anteriormente, sería interesante complementar este trabajo con otros de ámbitos sociales, culturales, económicos y políticos, para encontrar las razones que podrían estar influyendo en que la longevidad cambie debido a la ubicación geográfica del país. Seguidamente, entender las similitudes que en esos ámbitos tienen los países para poder agruparlos, al igual que se realizó en este estudio.

Eslovenia, a lo largo de los análisis tiende a confundirse entre el grupo del Oeste y el Este. Sus indicadores también reflejan esto, por lo que es un país que se debería tener más cuidado a la hora de situarlo en algún grupo. Se debería estudiar a Eslovenia en los siguientes años, para ver si su comportamiento tiende más a los países del Este o del Oeste y así definirlo más claramente.

Los países de la UE de los que no se disponía información, están situados en el Este de Europa, por lo que posiblemente, su comportamiento sea similar a los agrupados en esa zona. Pero se recomienda buscar esta información, para corroborar lo mencionado en cuanto a su similitud en comportamiento.

Por último señalar que este estudio ha sido realizado para el caso de los hombres pero también sería interesante realizarlo para el caso de las mujeres. De tal forma que se pudiese comparar si ambas casuísticas presentan similitudes o diferencias.

## Referencias

- BEZDEK, J. C., CORAY, C., GUNDERSON, R., Y WATSON, J. (1981). «Detection and characterization of cluster substructure I. Linear structure: fuzzy c-lines». *SIAM Journal on Applied Mathematics*, 40(2), 339-357.
- BILBAO, J. U. (2013). «El modelo social europeo: elementos constitutivos y viabilidad en un contexto adverso. Modelo social europeo: perspectivas de futuro frente a los desafíos económicos, geopolíticos, democráticos y demográficos del s.XXI». EUROBASK. ISBN: 978-84-695-9580-0.  
[http://eurobask.org/ficherosFTP/LIBROS/UNIVERSITAS\\_2013.pdf](http://eurobask.org/ficherosFTP/LIBROS/UNIVERSITAS_2013.pdf).
- BOOTH, H., HYNDMAN, R., TICKLE, L., Y DE JONG, P. (2006). «Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions». *Demographic Research*, 15, 289-310
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R., AND STONE, C. (1984). «Classification and Regression Trees». Wadsworth International Group, Belmont, California, USA.
- BRO, R., Y SMILDE, A. K. (2014). «Principal component analysis». *Analytical Methods*, 6(9), 2812-2831.
- CANUDAS-ROMO, V. (2008). «The modal age at death and the shifting mortality hypothesis». *Demographic Research*, 19, 1179-1204.

- CHRISTIANSEN, M. C., SPODAREV, E., Y UNSELD, V. (2015). «Differences in European mortality rates: A geometric approach on the age-period plane». *ASTIN Bulletin*, 1-26.
- CURRIE, I. D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4), 356-383.
- MEYER, D., DIMITRIADOU, E., HORNIK, K., WEINGESSEL, A. Y LEISCH, F. (2014). e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-4. <http://CRAN.R-project.org/package=e1071>
- DALSTRA, J., KUNST, A. E., GEURTS, J. J. M., FRENKEN, F. J. M., Y MACKENBACH, J. P. (2002). «Trends in socioeconomic health inequalities in the Netherlands, 1981–1999». *Journal of Epidemiology and Community Health*, 56(12), 927-934.
- DEBÓN, A., MARTÍNEZ-RUIZ, F., Y MONTES, F. (2012). «Temporal Evolution of Mortality Indicators: Application to Spanish Data». *North American Actuarial Journal*, 16(3), 364-377.
- DEBÓN, A., MONTES, F., Y SALA, R. (2005). «A comparison of parametric models for mortality graduation. Application to mortality data for the Valencia Region (Spain)». *SORT (Statistics and Operations Research Transactions)*, 29(2), 269-288.
- DEVESA, J. E. C., DEVESA, M. C., DOMÍNGUEZ, I. F., ENCINAS, B. G., MENEU, R. G. (2013). La sostenibilidad de los sistemas públicos de pensiones europeos como pilar fundamental de un nuevo Estado del Bienestar. Modelo social europeo: perspectivas de futuro frente a los desafíos económicos, geopolíticos, democráticos y demográficos del s.XXI. EUROBASK. ISBN:978-84-695-9580-0.[http://eurobask.org/ficherosFTP/LIBROS/UNIVERSITAS\\_2013.pdf](http://eurobask.org/ficherosFTP/LIBROS/UNIVERSITAS_2013.pdf).
- HUSSON, F. AND JOSSE, J. (2014). missMDA: Handling missing values with/in multivariate data analysis (principal component methods). R package version 1.7.3. <http://CRAN.R-project.org/package=missMDA>
- HUSSON, F., JOSSE, J., LE, S. AND MAZET, J. (2015). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 1.30. <http://CRAN.R-project.org/package=FactoMineR>
- FRIES, J. F. (1980). «Aging, natural death, and the compression of morbidity». *New England Journal of Medicine*, 303(3), 130-135.
- FRITZELL, J., KANGAS, O., BACCHUS HERTZMAN, J., BLOMGREN, J., Y HIILAMO, H. (2013). «Cross-temporal and cross-national poverty and mortality rates among developed countries». *Journal of Environmental and Public Health*, Volumen 2013, 1-15.
- GOERLICH, F. J., Y SOLER, Á. (2013). «Life potential as a basic demographic indicator». *Social Indicators Research*, 110(2), 537-548.
- HATZOPOULOS, P., Y HABERMAN, S. (2013). «Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data». *Insurance: Mathematics and Economics*, 52(2), 320-337.

- HUMAN MORTALITY DATABASE. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at [www.mortality.org](http://www.mortality.org) or [www.humanmortality.de](http://www.humanmortality.de) (Datos descargados Junio 2015).
- INSTITUTO NACIONAL DE ESTADÍSTICA (2013). Tablas de mortalidad Metodología, URL:<http://www.ine.es/jaxi/menu.do?type=pcaxisypath=%2Ft20%2Fp319ayfile=inebaseyL=0> [Consultado 03/07/2015].
- JACKSON, J. E. (2005). «A user's guide to principal components, volume 587». John Wiley y Sons.
- LEGUINA, J. (1973). «Fundamentos de demografía». Siglo XXI de España Editores, S.A. ISBN: 84-323-0089-6.
- LEON, D. A. (2011). «Trends in European life expectancy: a salutary view». *International Journal of Epidemiology*, 40(2), 271-277.
- LESCHKE, J. (2005). «Is it useful to cluster countries. Analysis of the example of unemployment insurance coverage of non-standard employed». Revised TLM.NET Conference Paper, Budapest. [http://www.siswo.uva.nl/tlm/confbuda/papers/papers\\_files/WP8Janine%20Leschke%20WZB%20revision%2011\\_2005.pdf](http://www.siswo.uva.nl/tlm/confbuda/papers/papers_files/WP8Janine%20Leschke%20WZB%20revision%2011_2005.pdf) [Consultado 03/07/2015].
- LIAW, A. AND WIENER, M. (2002). «Classification and Regression by randomForest». *R News* 2(3), 18-22.
- MACKENBACH, J. P. (2013). «Convergence and divergence of life expectancy in Europe: a centennial view». *European Journal of Epidemiology*, 28(3), 229-240.
- MACKENBACH, J. P., Y MCKEE, M. (2013). «A comparative analysis of health policy performance in 43 European countries». *The European Journal of Public Health*, 23(2), 195-201.
- MACKENBACH, J. P., KUNST, A. E., CAVELAARS, A. E., GROENHOF, F., GEURTS, J. J., Y EU WORKING GROUP ON SOCIOECONOMIC INEQUALITIES IN HEALTH. (1997). «Socioeconomic inequalities in morbidity and mortality in Western Europe». *The lancet*, 349(9066), 1655-1659.
- MESLÉ, F., Y VALLIN, J. (2002). «Mortality in Europe: the divergence between East and West». *Population*, 57(1), 171-212.
- MESLÉ, F. (2004). «Mortality in Central and Eastern Europe: long-term trends and recent upturns». *Demographic Research* 2, 45-70.
- O'DONNELL, O. (2009). «Measuring health inequalities in Europe». *Measuring and tackling health inequalities across Europe*, 15(3), 10.
- OLIDEN, P. E. (2009). «¿Existe vida más allá del SPSS? Descubre R». *Psicothema*, 21(4), 652- 655.
- ORALLO, J. H., QUINTANA, M. J. R., Y RAMÍREZ, C. F. (2004). «Introducción a la Minería de Datos». Pearson Prentice Hall.

- ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS) (2015). «Informe mundial sobre envejecimiento y salud». URL:<http://www.who.int/ageing/es/> [Consultado 27/01/2016].
- PAL, N. R., BEZDEK, J. C., Y HATHAWAY, R. J. (1996). «Sequential competitive learning and the fuzzy c-means clustering algorithms». *Neural Networks*, 9(5), 787-796.
- R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.
- REZAEI, M. R., LELIEVELDT, B. P., Y REIBER, J. H. (1998). «A new cluster validity index for the fuzzy c-mean». *Pattern recognition letters*, 19(3), 237-246.
- SHKOLNIKOV, V., ANDREEV, E., AND BEGUN, A. (2003). «Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples». *Demographic Research*, 8(11), 305–358.
- SPINAKIS, A., ANASTOSIOU, G., PANOUSIS, V., SPILIOPOULOS, K., PALAIOLOGOU, S., E YFANTOPOULOS, J. (2011). «Expert Review and Proposals for Measurement of Health Inequalities in the European Union: Full Report». European Commission.
- THERNEAU, T., ATKINSON, B. Y RIPLEY, B. (2015). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-9. <http://CRAN.R-project.org/package=rpart>
- ZAIDI, A. (2012). «Population ageing and financial and social sustainability challenges of pension systems in Europe: A cross-national perspective». *The Future of Multi-pillar Pensions*. Cambridge University Press: Cambridge, UK.