



Departamento de Economía
Programa de Doctorado en Economía Aplicada

TESIS DOCTORAL

Estudio de la mortalidad en España y sus provincias. Estimación e intervalos de confianza en subpoblaciones

Autor: Federico López Carrión

Directores:

Dr. José Luís Gutiérrez de Mesa
Dr. Luis Felipe Rivera Galicia

Alcalá de Henares, España
2014

*A ti, que te acercas a leerme.
A todo lo bueno de mi Patria, España.
A todos los que me quieren
y a todas las que me aman.*

El objeto del Gobierno es la felicidad
de la Nación, puesto que...
Constitución Española de 1812. Capítulo III. Artículo 13.

ÍNDICE

Capítulo 1.

| | |
|-----------------------------------|----------|
| Introducción | 1 |
| 1.1. Presentación | 3 |
| 1.2. Justificación del tema | 5 |
| 1.3. Objetivos de la Tesis | 7 |
| 1.4. Metodología | 10 |

Capítulo 2.

| | |
|---|-----------|
| La medición de la mortalidad | 13 |
| 2.1. Introducción..... | 15 |
| 2.2. Principios sobre la medición de la incidencia de la mortalidad | 15 |
| 2.3. Aproximación histórica | 22 |
| 2.4. La medición de la mortalidad en España..... | 26 |
| 2.5. Protocolo metodológico de Human MortalityDatabase | 33 |
| 2.6. Protocolo metodológico actual de la medición de la incidencia de la mortalidad en España | 44 |
| 2.7. Conclusiones y alternativas metodológicas | 51 |

Capítulo 3.

El análisis de la mortalidad en España y sus provincias en el período 1981-2008 65

| | |
|--|-----|
| 3.1. Introducción..... | 67 |
| 3.2. Tablas de mortalidad de España..... | 70 |
| 3.3. Tablas de mortalidad de las provincias españolas | 80 |
| 3.3.1. El riesgo de muerte | 81 |
| 3.3.2. La esperanza de vida | 88 |
| 3.4. Clasificación de las provincias por esperanza de vida | 100 |
| 3.4.1. Análisis por año | 103 |
| 3.4.2. Análisis combinado | 116 |
| 3.5. Conclusiones | 135 |

Capítulo 4.

Estudio sobre la proyección de la mortalidad en poblaciones de pequeño tamaño 141

| | |
|---|-----|
| 4.1. Introducción..... | 143 |
| 4.2. Fundamentos teóricos | 144 |
| 4.3. Aproximación histórica | 147 |
| 4.4. Proyección del riesgo de muerte provincial | 157 |
| 4.5. Proyección de la esperanza de vida provincial..... | 162 |
| 4.6. Resultados y conclusiones | 170 |

Capítulo 5.

Validación del intervalo de confianza del modelo lineal para la esperanza de vida provincial..... 173

| | |
|---|-----|
| 5.1. Introducción..... | 175 |
| 5.2. Esperanza de vida al nacimiento observada en 2009, 2010 y 2011..... | 176 |
| 5.3. Éxito del intervalo de confianza con la esperanza de vida observada..... | 179 |
| 5.4. Éxito del intervalo de confianza con las proyecciones a corto plazo..... | 183 |
| 5.5. Resultados y conclusiones | 196 |

Capítulo 6.

Conclusiones y líneas abiertas de investigación 199

| | |
|---|-----|
| 6.1. Introducción..... | 201 |
| 6.2. Difusión de la investigación | 201 |
| 6.3. Conclusiones | 204 |
| 6.4. Líneas abiertas de investigación | 211 |

Referencias..... 217

Capítulo 1

Introducción

- 1.1. Presentación
- 1.2. Justificación del tema
- 1.3. Objetivos de la tesis
- 1.4. Metodología

1.1. Presentación

En la lucha por la vida, interesa al hombre conocer el poder de la muerte. Científicos e investigadores prestan creciente atención a la mortalidad (véase, por ejemplo, Jones, Scheuerlein, Salguero-Gómez, Camarda, Schaible, Casper, Dahlgren, Ehrlen, García, Menges, Quintana-Ascencio, Caswell, Baudisch y Vaupel, 2014; Canudas-Romo y Becker, 2011; Canudas-Romo, 2010; Camarda, 2008; Girosi y King, 2008). La incidencia de la mortalidad tiene repercusiones en casi todos los ámbitos de la sociedad, incluyendo el Sistema Sanitario, el Sistema Educativo, el Ejército, los seguros de vida o el Sistema de Pensiones. El continuo descenso de la mortalidad y, por tanto, el aumento de la longevidad de la población tiene importantes consecuencias tanto a nivel individual como para la sociedad en su conjunto. Para la Demografía es esencial comprender la dinámica de la mortalidad con respecto a la edad y al tiempo, ya que este factor gobierna el cambio en las poblaciones.

La población es el sujeto y objeto de las acciones de progreso y desarrollo humano sostenible. Los datos sociodemográficos de los diversos grupos y sectores de la población según sea su edad y sexo cobran cada vez mayor relevancia y conforman la fuente de información que sirve para la planificación y gestión de la política pública, sabiendo además que la población no es solo un demandante de recursos, sino también un recurso productivo actor del mismo desarrollo. Por tanto, se convierte en prioridad para las Entidades Estadísticas la atención de la creciente y cada vez más cualificada demanda de datos, sobre las perspectivas de mortalidad de la población y sus características.

Las proyecciones de mortalidad tienen distintos usos prácticos. En primer lugar, se calculan periódicamente por los Organismos Internacionales como la ONU y las oficinas estadísticas nacionales, como parte integrante de las proyecciones de población. En segundo lugar, las agencias gubernamentales, como la Seguridad Social, el Ministerio de Economía, el Ministerio de Educación o el Ministerio de Defensa, confían en ellas cuando se asesoran en la toma de importantes decisiones estratégicas económicas, sociales y medioambientales que afectan a los ciudadanos. En tercer lugar

ejercen un gran influencia en el trabajo de los actuarios, profesionales de la rama de los seguros de vida y los grupos financieros.

España, constituido como Estado Social¹, defiende la igualdad de todos los ciudadanos dentro de su territorio. En este sentido, cabe recordar que las pensiones se distribuyen bajo unas reglas comunes que fijan la edad de jubilación a los 67 años. La razón de que se haya producido este retraso en la edad de jubilación es el incremento en la esperanza de vida que se observa a nivel nacional. La esperanza de vida al nacimiento² en 2012 fue 79,38 años para los varones y 85,13 años para las mujeres, 5,75 años superior. Con la tendencia actual de aumento de la esperanza de vida, un año cada lustro para las mujeres, una niña que nazca hoy llegará a cumplir los 102 años.

Paradójicamente, las reglas generales dan lugar a injusticias cuando observamos los casos particulares. La expectativa de vida depende de múltiples factores como el sexo, la edad, la genética, el estado civil, el ámbito geográfico, las condiciones sociales, la situación económica, el nivel educativo o la calidad medioambiental que disfruta cada individuo (véase, por ejemplo, Canudas-Romo, Gleit, Gómez Redondo, Coelho y Boe, 2008).

Ahora, supongamos el caso particular de un hombre que muere a los 67 años. La esperanza de vida para el total nacional es un índice general que olvida que era un varón soltero de una provincia del sur y de clase baja, analfabeto y residente en un barrio marginal afectado por altos niveles de contaminación.

Para facilitar el análisis territorial, el INE publica la esperanza de vida desagregada a nivel autonómico y provincial. En la España actual, una Comunidad Autónoma puede reclamar su esperanza de vida como hecho diferencial.

Teóricamente se puede calcular la esperanza de vida de una comunidad de vecinos. Otra cosa es la fiabilidad de los datos. Por ejemplo, la consistencia de la estimación de la esperanza de vida de las mujeres de Soria se base en una muestra de

¹ Según la Constitución Española de 1978.

² Fuente: INE. Tablas de mortalidad de la población de España 1991-2012 [www.ine.es].

46.271 individuos de las que no muere ninguna para 48 edades simples de las comprendidas entre 0 y 63 años observados³.

Hay abundante información sobre la mortalidad. Los indicadores son comúnmente aceptados aunque poco cuestionados y merecen una mayor atención. Existen razones sólidas para preguntarnos por la incidencia de la mortalidad y de la información especialmente relevante de sus índices, con especial atención a las poblaciones de pequeño tamaño.

1.2. Justificación del tema

Conocida la importancia que tiene la mortalidad en el campo de la planificación social y económica a nivel nacional y provincial interesa tender un nuevo puente, en la brecha que separa la Demografía y la Estadística, sobre el campo de la incidencia de la mortalidad en las poblaciones.

Mi experiencia profesional y mi estudio de la materia me han permitido detectar vacíos, poco reconocidos en la bibliografía, sobre la incidencia de la mortalidad en poblaciones pequeñas que es necesario cubrir. Ya hay importantes proyectos en marcha (véase, por ejemplo, Ahcan, Mevdev, Olivieri y Pitacco, 2014).

El enfoque de este trabajo se centra en un estudio integral de la mortalidad constituido por cuatro elementos: la medición, el análisis, el modelo de proyección y su validación.

En Estadística es imprescindible disponer de un patrón de medida fiable. Es necesario hacer una revisión conceptual sobre la medición de la incidencia de la mortalidad. Un protocolo metodológico consolidado, coherente y sistemático que facilite la comparación temporal y geográfica. Reconocidas instituciones han hecho importantes contribuciones, como el Human Mortality Database Protocol (Wilmoth, Andreev, Jdanvod y Glej, 2007), que aumentan con la creciente demanda de

³ Fuente: INE. MNP 2012. Defunciones por provincia de residencia, edad y sexo. [www.ine.es]

información sobre la incidencia de la mortalidad en los longevos. Véase, por ejemplo, (Gómez Redondo y García González, 2011; Guijarro y Peláez, 2008).

Por otra parte, la esperanza de vida al nacimiento es un indicador popular y ampliamente extendido, pero la medición de la incidencia de la mortalidad proporciona otros muchos indicadores biométricos. Todos estos indicadores merecen un estudio, pues sus propiedades pueden ser especialmente interesantes.

La medida de la incidencia de la mortalidad, de naturaleza contable, debe prestar una especial atención a la naturaleza estadística de su estimación, en especial en poblaciones de pequeño tamaño.

El segundo elemento, el análisis, toma sentido en base a la fiabilidad de la medición. La mortalidad es un fenómeno de las poblaciones y, sin embargo, se asocia a un territorio, tradicionalmente un país, sin que se observe especial interés sobre el hecho diferencial territorial. Un trato que descuenta importantes factores genéticos, geográficos, sociales, económicos o educativos, condicionantes de la mortalidad y que dan lugar a diferencias entre comunidades humanas.

Un análisis descriptivo nacional permite apreciar la evolución y muestra los patrones de la mortalidad en la población en su conjunto. Por otra parte, un análisis provincial descubre la distribución geográfica de la mortalidad, permite clasificar las provincias de acuerdo a sus particularidades y proporciona argumentos para mejorar las proyecciones. Adicionalmente, la visión provincial centra la atención en un factor crítico, como es el comportamiento de la incidencia de la mortalidad en poblaciones de pequeño tamaño. El descenso continuo de la mortalidad deja expuestas a las poblaciones frente a factores aleatorios, originando mayores fluctuaciones y una creciente cantidad de edades en las que no se observan defunciones, que complican la estimación de la incidencia de la mortalidad.

El tercer elemento del estudio aspira a una ambición universal: conocer el futuro. Analizando la evolución observada podemos especificar un modelo de proyección de la mortalidad provincial que describa el comportamiento observado y satisfaga las necesidades de los usuarios. Existe un pujante interés en esta línea de

investigación. Véase, por ejemplo, French (2014), Wilmoth, Zureick, Canudas-Romo, Inoue y Sawyer (2012) o Guijarro y Peláez (2009).

Y por último, pero no menos importante, el cuarto elemento esta constituido por la validación del modelo de proyección de la mortalidad provincial frente a datos observados y la comparación con otras proyecciones. Es necesario evaluar la eficacia y la eficiencia del modelo y, sobre todo, la utilidad para los interesados, verdaderos destinatarios del producto.

Es necesario por tanto estudiar en qué medida el razonable interés por conocer la singularidad de la esperanza de vida de un individuo es compatible con la necesidad de disponer de un tamaño poblacional lo suficientemente grande para que la estimación sea consistente. Reconocer el problema y aceptar la situación como inevitable e intrínseca al tamaño de la población y buscar un modelo que dé respuesta a la demanda de proyecciones futuras para cualquier población y en particular para las poblaciones que, aun siendo de pequeño tamaño, por razones administrativas o de otro tipo, es necesario satisfacer.

1.3 Objetivos de la tesis

Esta Tesis se marca como principal objetivo cubrir los vacíos detectados a lo largo de mi investigación de la incidencia de la mortalidad en poblaciones, con especial atención a las poblaciones de pequeño tamaño. Con este objetivo, la Tesis acomete un estudio integral concentrado en los cuatro elementos capitales: la medición y el análisis de la mortalidad, la especificación del modelo y su validación.

Para completar este estudio, se diseña un plan estratégico que aborda los elementos capitales implicados en el estudio de la mortalidad.

El capítulo 2 revisa los principios de la medición de la incidencia de la mortalidad. Se presenta el diagrama de Lexis y la tabla de mortalidad como herramientas fundamentales para la medición de la incidencia de la mortalidad. A lo largo de la Historia, los investigadores se han preocupado por establecer una metodología de contabilidad integrada en un marco sistemático y completo que

facilite la comparación temporal y geográfica de la mortalidad. En los apartados 2.3, 2.4 y 2.5 se realiza una revisión histórica que describe la evolución de la medición de la mortalidad que culmina con el protocolo metodológico del HMD. En el apartado 2.6 se detalla el actual protocolo metodológico utilizado en España y en el apartado 2.7 se incluyen las conclusiones y se proponen alternativas metodológicas para mejorar la medición de la incidencia de la mortalidad en España. Además se aborda el tema de cuál debería ser el tamaño mínimo de una población para que la estimación de la incidencia de la mortalidad sea consistente.

El capítulo 3 utiliza la metodología del INE (2009) para elaborar las tablas de mortalidad de España desde 1981 hasta 2008 y sus provincias desde 1991 hasta 2008, obteniendo 1.480.064 datos sobre la incidencia de la mortalidad. El apartado 3.2 incluye un análisis descriptivo de la mortalidad en España, a lo largo del tiempo y por edad, completo, exhaustivo y detallado.

El análisis de la mortalidad provincial se trata en el apartado 3.3. La primera parte de este epígrafe se centra en el riesgo de muerte, presentando distintos gráficos explicativos sobre la elevada fluctuación de las series temporales así como destacando la correlación de la fluctuación con el tamaño de la variable. Se completa con un conjunto de mapas en los que se puede apreciar la distribución geográfica del riesgo de muerte. A continuación sigue un análisis descriptivo de la esperanza de vida provincial, en el apartado 3.3.2, con gráficos de las series y mapas en los que se muestra la distribución geográfica de la esperanza de vida. La comparación entre los dos indicadores de la intensidad de la mortalidad permite apreciar la mayor estabilidad temporal y la contigüidad geográfica de la esperanza de vida.

En el apartado 3.4, aprovechando la mayor estabilidad de la esperanza de vida, se aborda una clasificación de las provincias. Se utiliza un análisis de componentes principales con dos enfoques: Año por año, que se aborda en el apartado 3.4.1 y un segundo enfoque con un particular énfasis en el análisis combinado que resume la evolución de las provincias a lo largo del tiempo en un mismo plano. Un dendrograma de clasificación completa la discriminación provincial. El resultado del capítulo es una interesante clasificación de las provincias en España atendiendo a su mortalidad.

De acuerdo con el análisis descriptivo realizado, el capítulo 4 propone dos novedosos modelos para la proyección de la mortalidad provincial. Los modelos se basan en una metodología relacional y, además de una estimación puntual, proporcionan una estimación por intervalos que sirve como medida de la bondad de la predicción. En el apartado 4.4 se especifica el modelo para el riesgo de muerte que resuelve el problema de la inclusión del cero en el soporte de la variable con un modelo probabilístico mixto. En el apartado 4.5 se especifica un modelo para la proyección de la esperanza de vida provincial con resultados altamente satisfactorios.

El capítulo 5 se ocupa de evaluar el modelo de proyección propuesto para la esperanza de vida provincial. En el apartado 5.2 se presentan y analizan los resultados de las tablas de mortalidad observadas en los años 2009 a 2011. En el apartado 5.3 se validan frente a los datos observados y en el apartado 5.4 se comparan con las proyecciones de mortalidad a corto plazo que elabora el INE (2008).

El capítulo 6 reúne la difusión de mis investigaciones previas hasta la fecha, así como las conclusiones y resultados de esta Tesis y las líneas abiertas de investigación que me propongo continuar en el futuro junto con los investigadores que estén interesados.

En resumen, esta Tesis ha demostrado la importancia de la dimensión geográfica en la incidencia de la mortalidad. Se han propuesto dos novedosos modelos para la proyección de la mortalidad provincial. Un modelo para el riesgo de muerte que salva el problema del aumento de la frecuencia observada de ceros y otro modelo que proporciona, además de una estimación puntual, una estimación por intervalos eficaz, eficiente y útil.

Ambos modelos se pueden aplicar a cualquier subpoblación incluida en una población de referencia con la única condición de que la forma de medir la incidencia de la mortalidad sea la misma.

1.4 Metodología

El proceso metodológico para la elaboración de la presente Tesis se ha basado en un ejercicio continuo de ensayo y error. El resultado ha sido un sinfín de fracasos entre los que aparecen también importantes éxitos que se recogen como resultados en este documento. En cuanto a los fracasos, también merecen su reconocimiento pues han ayudado a consolidar el trabajo y permiten valorar en su justa medida los frutos obtenidos.

Esta investigación ha requerido una amplia revisión personal de bibliografía y el análisis crítico de una pluralidad de textos relacionados con la mortalidad y todo el entorno histórico, internacional y metodológico de su medición. Junto al material de documentación han jugado un papel clave las aportaciones de distintos profesionales del ramo de la Demografía y la Estadística. En primer lugar mis directores de Tesis y compañeros en este camino, José Luíz Gutiérrez de Mesa y Luíz Felipe Rivera Galicia. Me han animado y han participado decisivamente en el estudio. Me han iluminado y han sabido reconducirme cuando estaba empeñándome en la dirección equivocada. No hay gloria sin dolor, es lo que tiene.

La discusión con mis compañeros del Área de Análisis y Proyecciones Demográficas en el Instituto Nacional de Estadística ha enriquecido el estudio, Margarita Cantalapiedra, desde su espíritu crítico, o Isabel González y el resto del equipo, suministrando información. Ignacio Arbués y David Salgado, de la Unidad de Apoyo a la Dirección General de Metodología, Calidad y Tecnología de la Información y las Comunicaciones, han ayudado con aportaciones teóricas y en la implementación del código para la programación en R de las distintas rutinas empleadas al efecto. Desde la distancia, Carlo di Camarda, en su momento en el Instituto Max Planck para Investigaciones Demográficas de Rostock y actualmente en el Instituto de Estudios Demográficos Francés de París, me ha proporcionado ayuda en la implementación de código para la elaboración de superficies en lenguaje R. Otras personas de la Subdirección de Estadísticas de Población del INE que han colaborado son Mayte Gutiérrez, Iciar Caballero, Sara Merino, Pepa Epa, Víctor López y Miguel Macarro y

desde otras unidades en el INE, María Luisa Calvo y Marce García. Miguel Guigó, Estadístico Facultativo del INE, actualmente desplazado a Eurostat en Luxemburgo, me puso en contacto con mis directores de Tesis y Tomás López, mi padre y profesor de Estadística en la Facultad de Económicas y Empresariales de la Universidad de Valladolid, es el promotor original de esta Tesis.

Junto con libros, revistas científicas y documentación oficial, se han utilizado también como herramientas metodológicas con las que procesar los conjuntos de datos más voluminosos, el paquete estadístico SAS. El lenguaje R se ha utilizado en el desarrollo de los algoritmos estadísticamente más técnicos y la hoja de cálculo Excel, sobre todo en la presentación de los resultados. También ha sido útil Internet en la búsqueda sobre toda clase de información, documentación y procedimientos estadísticos.

La planificación estratégica del trabajo se ha guiado por el principio de responder a los objetivos de la Tesis abordando los elementos capitales: medición, análisis descriptivo, modelización y validación. De ellos se derivan los resultados y conclusiones que se presentan en este volumen.

El proceso de construcción de esta Tesis ha sido análogo al de una cadena. El primer eslabón y embrión fue el Trabajo de Investigación Tutelado que se ocupó de la medición de la incidencia de la mortalidad y que se ha transformado en el segundo capítulo. En base a este primer eslabón se engarzó, calculando las tablas de mortalidad que proporcionaron los datos para el análisis descriptivo, el segundo eslabón, y que tomó forma en el tercer capítulo. El tercer eslabón es la especificación de los modelos de proyección provincial que se trata en el capítulo cuarto para pasar en el quinto capítulo a la validación del modelo que se incorpora como un cuarto eslabón. Finalmente se completa la cadena con el capítulo sexto a modo de broche y este primer capítulo como estuche de presentación.

Durante todo el proceso metodológico se ha ido dando forma a cada capítulo y apartado incorporando nuevos elementos que completaban las distintas partes del cuerpo de esta Tesis hasta pulir todos los detalles. Y a lo largo del trabajo desarrollado se han descubierto interesantes líneas abiertas de investigación.

Con la ayuda del plan estratégico y perseverando en el trabajo, se han alcanzado los objetivos que se fijaron para dar finalmente a la Tesis la forma definitiva.

La presente Tesis, más allá de constituir un circuito cerrado de aportaciones, aspira a abrir líneas de investigación en el estudio de la mortalidad y despertar la inquietud de los investigadores por novedosos temas que diversifiquen la red del conocimiento humano sobre la mortalidad en las poblaciones, que es tanto como decir de la vida.

Capítulo 2

La Medición de la Mortalidad

- 2.1. Introducción
- 2.2. Principios sobre la medición de la incidencia de la mortalidad
- 2.3. Aproximación Histórica
- 2.4. Medición de la mortalidad en España
- 2.5. Protocolo metodológico de Human Mortality Database
- 2.6. Protocolo metodológico actual de la medición de la incidencia de la mortalidad en España
- 2.7. Conclusiones y Alternativas metodológicas

2.1. Introducción

Esta sección ofrece las principales herramientas y fundamentos para el estudio de la mortalidad. El lector interesado en ampliar sus conocimientos metodológicos sobre la construcción de tablas de mortalidad puede consultar los manuales clásicos de Keyfitz (1979) y Preston, Heuveline y Guillot (2001), o una aproximación más práctica en Wilmoth, Andreev, Jdanov y Gleit (2007).

Este capítulo se desarrolla en 5 apartados.

- En el segundo apartado se introducen los principios básicos para la medición de la incidencia de la mortalidad.
- El tercero se ocupa de un recorrido histórico, antecedentes e hitos del estudio de la mortalidad, así como una descripción del estado actual del arte.
- El cuarto consiste en una aproximación al caso particular del estado actual del protocolo metodológico oficial para medir la incidencia de la mortalidad en España.
- Se termina con un quinto apartado donde se recogen los resultados del capítulo y se presentan distintas alternativas metodológicas que mejoran importantes aspectos de la medición de la incidencia de la mortalidad en España.

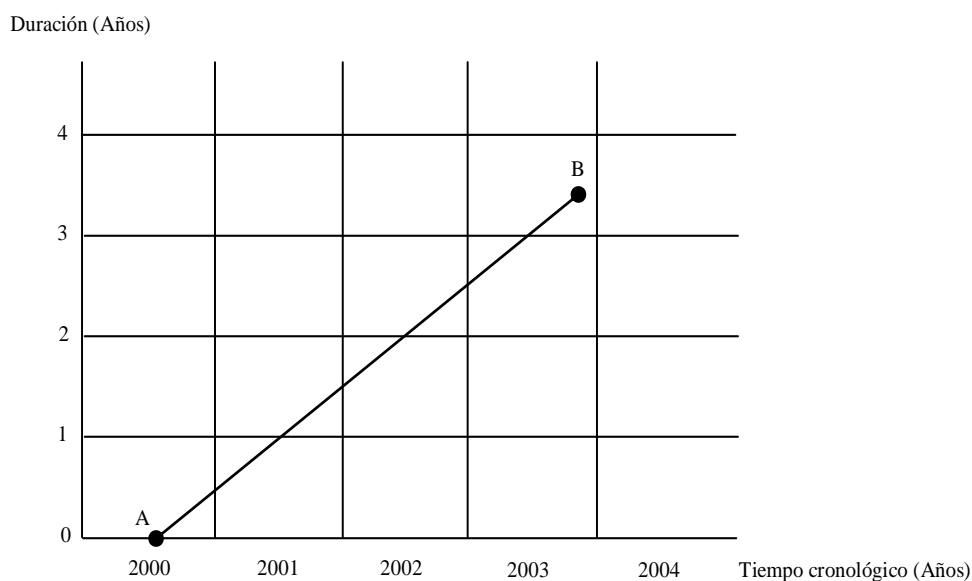
2.2. Principios sobre la medición de la incidencia de la mortalidad

El diagrama de Lexis es una herramienta gráfica para el estudio del fenómeno demográfico de la mortalidad. Consiste en un plano cartesiano donde el eje de abscisas marca la fecha cronológica y el eje de ordenadas marca la edad. Cada individuo de una población tiene en cada fecha cronológica t una edad x . Una línea de vida es un segmento rectilíneo que representa la vida de un individuo. Es paralela a la

bisectriz del primer cuadrante y tiene un extremo en el eje de abscisas que representa la fecha de nacimiento.

En el diagrama de Lexis de la Figura 2.1 se puede observar la línea de vida de un individuo que nace a mediados del año 2000 (punto A) y muere a finales del año 2003 (punto B).

Figura 2.1. Diagrama de Lexis con línea de vida desde el punto A, nacimiento, al punto B, defunción.

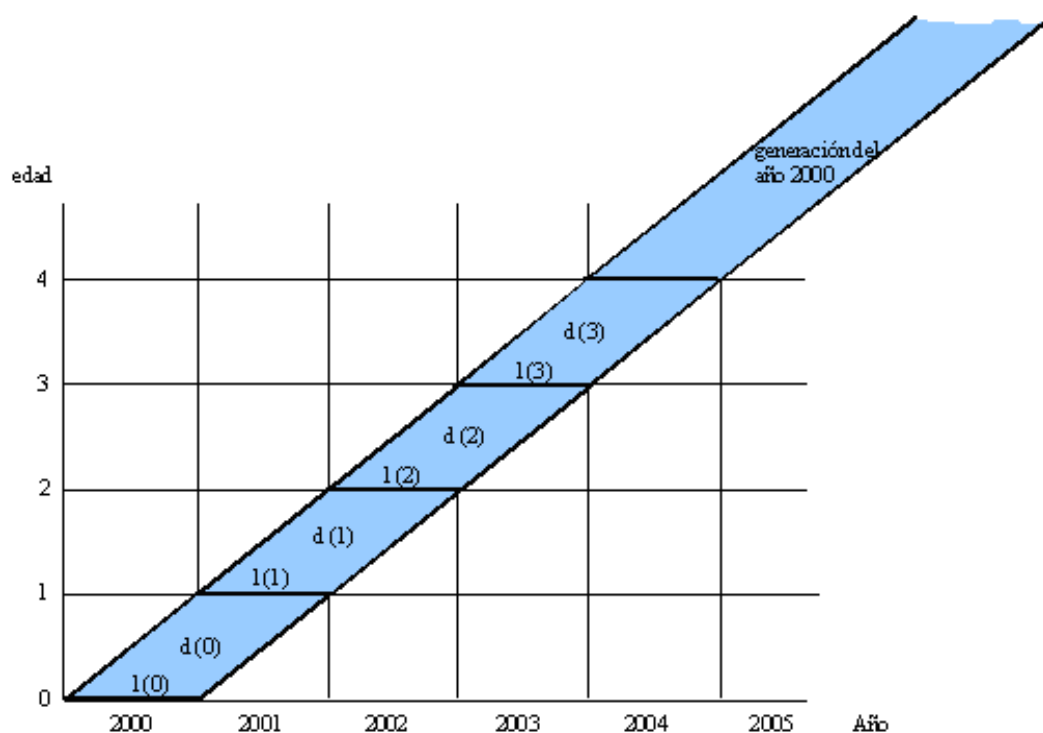


Fuente: Elaboración propia

Una generación es un conjunto de individuos que nacen el mismo año. En un diagrama de Lexis la generación del año t se corresponde con las líneas de vida incluidas en la región abierta del plano limitado por el eje de abscisas y las paralelas a las líneas de vida que pasan por los puntos del eje de abscisas que representan el 1 de enero del año t y del año $t+1$.

Por ejemplo la generación del 2000 está formada por todas las líneas de vida que parten del segmento del eje de abscisas del año 2000 y que, por tanto, están incluidas en la región abierta coloreada de la Figura 2.2.

Figura 2.2. Diagrama de Lexis representando la generación del 2000.



Fuente: Elaboración propia

Si $l(x)$ representa el número de supervivientes con x años, y $d(x)$, el número de defunciones de individuos con x años se tiene que, de acuerdo con el diagrama de Lexis de la Figura 2.2 el total de nacidos vivos en el año 2000, $l(0)$, esta representado por el número de líneas de vida que parten del segmento del eje de abscisas para el año 2000, las defunciones de la generación con 0 años, $d(0)$, son el número de líneas de vida que terminan en el paralelogramo determinado por la región oscura coloreada, en el diagrama de Lexis superior, y limitada por las líneas de aniversario 0 y 1. El número de supervivientes con 1 año, $l(1)$, es el total de líneas de vida que cruzan la línea de aniversario 1, las defunciones con 1 año, $d(1)$, es el total de líneas de vida que terminan en la región oscura coloreada limitada por las líneas de aniversario 1 y 2. Los restantes valores de la Figura [$l(2)$, $d(2)$, $l(3)$ y $d(3)$] se calculan de forma análoga según el procedimiento anteriormente descrito.

Las tablas de mortalidad, también denominadas tablas de vida¹, constituyen un modelo fundamental para representar numéricamente el modo en que una población experimenta el cambio de estado de sus individuos a lo largo del tiempo. Un proceso estocástico a manera de cadena de Markov donde los estados son las edades 0, 1, 2, etc. y un estado extra de fallecido, -1, junto con las probabilidades de transición de estado a estado (Hoem, 1977).

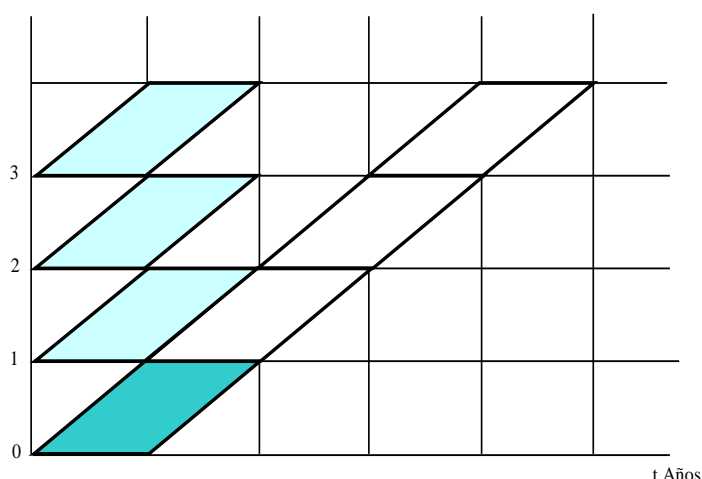
Conceptualmente hay dos formas principales de expresar las tablas de mortalidad:

- La tabla longitudinal o de generación. En el diagrama de Lexis de la Figura 2.2 la tabla longitudinal del año 2000 se corresponde con la región abierta oscura coloreada que se prolonga hasta la extinción de todos los individuos que nacieron en el año 2000. Requiere el seguimiento de todos los individuos de la generación objeto de estudio a lo largo del tiempo. Aunque en teoría es la forma más correcta de estudiar la mortalidad por observar exclusivamente los individuos sujetos al riesgo de morir, presenta grandes inconvenientes. En primer lugar, el periodo de recogida de información se extiende por muchos años –más de un siglo seguramente– con lo que las fuentes de información disponibles son menos fiables. Por otra parte es prácticamente imposible garantizar que la población objeto de estudio sea un ente cerrado y aislado de las migraciones. A esto hay que añadir que el interés por un colectivo humano que ha desaparecido es relativo, y más bien histórico.
- Tabla transversal. Trata de evitar los inconvenientes citados de las longitudinales. Una tabla de mortalidad de tipo transversal estudia como incide la mortalidad en la población en un momento o periodo determinado, que incluirá muchas generaciones. A partir de la observación transversal del momento, se construye una generación ficticia para la que se estiman todas las funciones de la tabla de mortalidad. En la Figura 2.3 se puede observar

¹ Life tables en el original inglés. Una visión anglosajona más optimista frente a nuestra versión latina, tabla de mortalidad, como apunta Goerlich (2008).

como existe una correspondencia entre cada uno de los paralelogramos que constituyen la generación real y los paralelogramos de la generación ficticia que se construye. A pesar de no tratarse de una generación real, esta aproximación es aceptada en general sin olvidar que la incidencia de la mortalidad describe los cambios de estado, de vivo a muerto, de un individuo teórico. Un individuo que se incorpora a la cohorte en ese periodo queda afectado por la incidencia de la mortalidad en distinto grado según pasen los años al mismo tiempo que va cambiando la intensidad de la mortalidad para cada edad. Es interesante notar que de esta forma se concibe la población de la cohorte ficticia como un ente global, un organismo singular frente al tradicional concepto de agregación de individuos, atribuyéndole personalidad propia al conjunto de la población en un periodo concreto de tiempo.

Figura 2.3. Diagrama de Lexis con observación transversal más oscura.



Fuente: Elaboración propia

A continuación se presentan las funciones biométricas de una tabla de mortalidad completa, i.e., por edades simples en la que x recorre los números naturales. La cota superior del conjunto de las edades ha ido subiendo con el paso del tiempo. Actualmente suele cerrar a los 100 años aunque en algunos casos se acorta o

se prolonga la serie en función de la disponibilidad de efectivos suficientes y atendiendo a la creciente demanda de información sobre *superlongevos*².

La función **supervivientes con x años**, $l(x)$, es el número de personas de una generación ficticia, expuesta a la intensidad de la mortalidad de la población objeto de estudio, que llegan a cumplir los x años. Es una variable cuya magnitud se mide en personas.

La raíz de la función de supervivientes o primer valor $l(0)$ es, por convenio, 100.000. Quizá la razón principal sea de carácter empírico pues se observa que para poblaciones de tamaño menor de 100.000 la distribución de probabilidad de sucesos raros, i.e., de probabilidad pequeña como el fallecimiento de un individuo, distorsiona fuertemente los resultados de la tabla de mortalidad.

Otras razones son de carácter práctico. En caso de tomar un valor menor, por ejemplo de 100, i.e., expresar la función de supervivientes en tanto por ciento, la función de defunciones sólo sería un entero a partir de riesgos de mortalidad del orden de las centésimas. Esta condición solo se cumple en las tablas de mortalidad completas actuales para edades superiores a los 40 años. Por otra parte 1.000.000 sería excesivo. La población objeto de estudio suele ser de menor tamaño. En el caso de España, no más de 14 provincias están claramente por encima de esta cifra.

La función **defunciones a los x años**, $d(x)$, es el número de individuos de la generación que fallecen con x años de edad. Se mide en personas y verifica que:

$$d(x) = l(x) - l(x + I) \quad (2.1)$$

La función **tiempo vivido el último año de vida**, $a(x)$, es la media del tiempo vivido por los individuos que fallecen con x años desde su último cumpleaños. Se mide en años.

La función **probabilidad o riesgo de muerte con x años**, $q(x)$, es la probabilidad de que una persona perteneciente a la generación y que haya llegado a

² El lector interesado en el campo de los superlongevos puede consultar la entrada “superlongevos” del blog del demógrafo Julio Pérez Díaz (CSIC) <http://apuntesdedemografia.wordpress.com/envejecimiento-demografico/longevidad-y-limites-del-envejecimiento-demografico/los-superlongevos/>

cumplir los x años, fallezca con x años de edad. Es una magnitud adimensional y se define como:

$$q(x) = \frac{d(x)}{l(x)} \quad (2.2)$$

La función **población estacionaria con x años**, $L(x)$, es el total de años vividos por la generación entre las edades x y $x+1$. Se mide en años-persona³ y se define como:

$$L(x) = l(x+1) + a(x) \cdot d(x) \quad (2.3)$$

La población estacionaria representa la estructura por edades que tendría una población a final de año sometida a las mismas condiciones de mortalidad de la población real, descrita por los riesgos de muerte dados en la tabla de mortalidad, donde se incorporen todos los años 100.000 efectivos nuevos.

La función **tasa de mortalidad con x años**, $m(x)$, es un cociente donde el numerador son las defunciones y el denominador la exposición al riesgo. Es una magnitud adimensional y se calcula como:

$$m(x) = \frac{d(x)}{L(x)} \quad (2.4)$$

La función **total de años vividos desde los x años**, $T(x)$, es el tiempo que le queda por vivir a la generación desde los x años de edad hasta su extinción. Se mide en años-persona y se define como:

$$T(x) = T(x+1) + L(x) \quad (2.5)$$

La función **esperanza de vida con x años**, $e(x)$, es el número promedio de años que le quedan por vivir a un individuo de la generación que ha cumplido los x años. Se mide en años y se expresa como:

³ Análogamente a los años-luz, que miden una magnitud como tiempo por velocidad, la población estacionaria es una magnitud que se define como producto de tiempo por fallecidos y que se mide en años-persona.

$$e(x) = \frac{T(x)}{l(x)} \quad (2.6)$$

Dada una cualquiera de las funciones $l(x)$, $d(x)$ ó $q(x)$, el resto queda determinada. Si además se tiene $a(x)$ el resto de las funciones de la tabla de mortalidad, $m(x)$, $L(x)$, $T(x)$ y $e(x)$, también quedan determinadas unívocamente. Con dos funciones de la tabla, una del primer grupo y otra del segundo, por ejemplo a partir de $e(x)$ y $a(x)$, se puede reconstruir el resto de la tabla de mortalidad.

Desarrollando la ecuación 2.6, se tiene:

$$e(x) = \frac{T(x)}{l(x)} = \frac{T(x+1) + L(x)}{l(x)} = \frac{e(x+1) \cdot l(x+1) + l(x+1) + a(x) \cdot d(x)}{l(x)} \quad (2.7)$$

$$e(x) \cdot l(x) = e(x+1) \cdot l(x+1) + l(x+1) + a(x)[l(x) - l(x+1)] \quad (2.8)$$

Dada la serie de $e(x)$ se reconstruye la serie de supervivientes $l(x)$ recurrentemente despejando de la ecuación 2.7.

A partir de $l(0) = 100.000$ se obtiene $l(1)$ en función de $a(0)$ y en general $l(x+1) = l(a(x))$.

2.3. Aproximación Histórica

No ha sido corto ni fácil el camino recorrido en este sector de la estadística vital. El interés por el análisis de las cuestiones demográficas viene de lejos. Ya el libro de los Números del Antiguo Testamento consigna con minuciosidad extrema: la cantidad de jefes de las tribus (cap. 7), número de las poblaciones y libaciones necesarias (13) o la cantidad de hombres sublevados (16:2)⁴.

⁴ Además de cabezas de ganado que han de ser destinadas al sacrificio ritual (28-29), cantidad de botín y su reparto exacto (31), agrimensura y dimensiones del territorio (35), etcétera.

Las tablas de mortalidad son desde hace tiempo una herramienta de bioestadísticos y actuarios para estudiar la incidencia de la mortalidad. Este instrumento estadístico básico es además un artefacto histórico. Ha sido forjado y refinado por una serie de matemáticos a lo largo de los siglos XVII y XVIII con dos periodos sobresalientes de creatividad en torno a 1670 y 1740 (Dupâquier, 1996). Fueron matemáticos y astrónomos los que desarrollaron las tablas de mortalidad y se merecen ser consideradas como uno de los logros punteros de la revolución científica. Los historiadores de la Ciencia señalan la matematización del conocimiento en los siglos XVII y XVIII y a menudo centran el interés en el cálculo y sus aplicaciones a la Física. Olvidan así la importancia del desarrollo simultáneo de la aplicación del cálculo y la incipiente probabilidad en las Ciencias Sociales y Humanas. La disciplina experimentó un progreso discontinuo y caótico a través de los refinamientos matemáticos y estadísticos que animaron el progreso de la Demografía.

Sobre todos los factores científicos, los pioneros estuvieron condicionados por la falta de fuentes de información con datos fiables. El primer país en disponer de información suficientemente buena para construir tablas de mortalidad es Suecia en 1751⁵. Los registros nacionales empezaron a consolidarse a partir del siglo XIX, en Francia (1816) y Dinamarca (1835).

El cálculo de tablas de mortalidad está estrechamente asociado a las fuentes de información disponibles. Estas carencias no fueron obstáculo para que algunos valientes lucharan contra la adversidad. La obra de Graunt (1662) aprovechó un incipiente registro de nacimientos y defunciones y dio como resultado un trabajo basado en datos de la población de Londres entre 1601 y 1661.

⁵ Dato extraído de Human Mortality Database, proyecto internacional cuyo objetivo es recoger la mayor cantidad de datos existentes y establecer un marco común para la comparación de la mortalidad entre países y a lo largo del tiempo.

Tabla 2.1. Serie de supervivientes calculada por Graunt.

| Edad | $l(x)$ |
|------|--------|
| 0 | 100 |
| 6 | 64 |
| 16 | 40 |
| 26 | 25 |
| 36 | 16 |
| 46 | 10 |
| 56 | 6 |
| 66 | 3 |
| 76 | 1 |

Fuente: Elaboración propia a partir de datos de Graunt (1662)

Las fuentes de las que disponía Graunt indicaban el sexo y la causa de muerte, pero no la edad. Así que estimó los supervivientes para cada edad. Por ejemplo, para estimar los individuos que morían antes de los 6 años tuvo en cuenta la proporción de personas fallecidas de enfermedades infantiles y añadió la mitad de las que fallecían por sarampión o viruela cuyo resultado era 36%.

Los trabajos de Graunt y Edmund Halley (1693), demógrafo además de astrónomo, sobre la estimación de los grados de mortalidad humana basado en “curiosos estados de nacimientos y defunciones de la ciudad de Breslau durante los años entre 1687 y 1691” constituyeron, en el terreno metodológico, un importante empuje en la construcción de las tablas de mortalidad. Halley señaló los errores de Graunt e ideó su propio método que, mal interpretado unas veces y ligeramente mejorado otras, fue durante más de siglo y medio el único método corriente si se exceptúa la innovación que Wargentin (1766) incorporó en las tablas para Suecia con cifras de defunciones y de censo de población.

En un principio, las tablas de mortalidad estaban formadas por, a lo sumo, dos series de datos: los números de fallecidos y de supervivientes de las distintas edades. Posteriormente se fueron incorporando la probabilidad de muerte o la esperanza de vida.

Milne (1815) publicó su primer trabajo, basado tanto en la información de la población como de los fallecidos clasificados por edad y sexo, que se puede considerar tabla de mortalidad para la ciudad de Carlisle, Inglaterra con datos del periodo 1779 a 1787. Una tabla de mortalidad consolidada, pues estaba basada no

solo en una muestra de defunciones por edad sino que también incorporó la población por edad expuesta al riesgo.

Mientras, la inquietud por la Demografía crecía. En 1798 Malthus publicó su libro más famoso, *An Essay on the Principle of Population*. En el capítulo 1 escribe su apocalíptico augurio sobre el crecimiento de la población:

*"The power of population is indefinitely greater than the power in the earth to produce subsistence for man"*⁶

Acertó en la evolución del tamaño poblacional, pero afortunadamente dejó de valorar en su justa medida la capacidad tecnológica del hombre para aumentar los rendimientos del trabajo.

En 1825 el matemático y actuario británico Benjamin Gompertz elaboró una nueva serie de tablas de mortalidad para la Royal Society. Este trabajo le sugirió su ley de la mortalidad humana.

El periodo entre 1860 y 1910 es un periodo de transición en el que la demografía surge de la estadística como un campo de interés separado con entidad propia. Durante estos años aparecen grandes demógrafos como Adolphe Quételet (1796–1874), William Farr (1807–1883), Louis-Adolphe Bertillon (1821–1883) y su hijo Jacques (1851–1922), Joseph Körösi (1844–1906), Anders Nicolas Kaier (1838–1919), Richard Böckh (1824–1907), Émile Durkheim (1858-1917), Wilhelm Lexis (1837–1914) y Luigi Bodio (1840–1920) . Todos ellos contribuyeron al desarrollo de la demografía y del conjunto de herramientas y métodos del análisis demográfico en general y de la mortalidad en particular.

Desde entonces el progreso de la ciencia de la Demografía en general y del estudio de la mortalidad en particular ha experimentado un notable avance. A partir de mediados del siglo XX la escuela francesa dominó la materia de la mano de ilustres

⁶ La traducción es "El poder de la población es indefinidamente mayor que el poder en la Tierra de producir subsistencias para el hombre". Malthus T.R. 1798. *An Essay on the Principle of Population*.

nombres como Pressat, Henry o Sauvy. En los años 70, con los matemáticos Jacques Vallin, Keyfitz, Preston y Thatcher se consolidan los conceptos y definiciones en el estudio del fenómeno de la mortalidad. La medición de la mortalidad se convierte en un asunto contable.

Así como para el cálculo de las Cuentas Nacionales hay un reglamento de la ONU que establece las directrices coherentes para el cálculo de magnitudes macroeconómicas comparables tanto espacialmente entre los distintos países como temporalmente a lo largo del tiempo, en demografía ningún organismo internacional ha establecido criterios comunes para la elaboración de las tablas de mortalidad. Cada país utiliza su propia metodología, lo que dificulta las comparaciones internacionales. El organismo estadístico encargado en cada país considera los supuestos más adecuados condicionados en gran medida por las fuentes de información.

2.4. La medición de la mortalidad en España

En España, según Leguina (1989), los primeros datos de mortalidad por edades que se conocen se refieren a 1860 y fueron publicados por Miguel Merino en 1866. El Instituto Geográfico y Estadístico elaboró, con grandes dificultades debido a la falta de datos, una segunda tabla de mortalidad referida a 1880 con datos de fallecidos entre 1878 y 1882.

El cálculo de tablas de mortalidad constituye una tarea de larga tradición en el INE. En cada ocasión se han llevado a cabo a partir de los desarrollos metodológicos considerados como más pertinentes, condicionado a la disponibilidad de las informaciones de base sobre defunciones y referentes poblacionales disponibles.

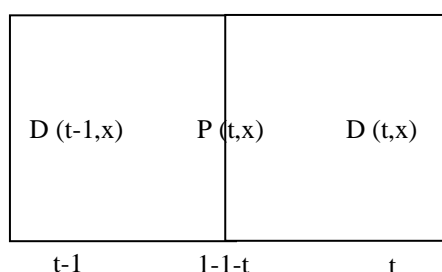
En 1946 el INE publica la primera tabla de mortalidad de la Población España relativa al bienio 1930-31 (INE, 1946). Es una tabla calculada a partir de las defunciones de dos años y referida a 31 de diciembre de 1930 aprovechando el censo de población a esa fecha con datos por edad y sexo. La fuente de información para las defunciones es el MNP registrados por sexo y edad y no cuenta con anotaciones por

generación. Los datos presentan errores como consecuencia de la tendencia a declarar edades terminadas en 0 y 5.

La serie de defunciones y poblaciones se suaviza con interpolaciones para corregir la concentración indebida de las cifras en determinadas edades.

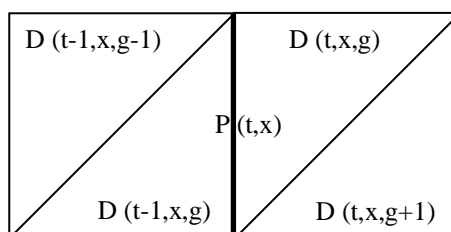
El riesgo de mortalidad, punto de partida de la tabla de mortalidad, se estima de acuerdo con la hipótesis de que las defunciones se distribuyen uniformemente por año y edad con lo que para las dos generaciones de cada año t y edad x , $D(t,x,g)$ y $D(t,x,g+1)$ de la Figura 2.6 valen lo mismo.

Figura 2.5. Representación de la región con año $t-1$ y t y edad x del diagrama de Lexis.



Fuente: Elaboración propia

Figura 2.6. Representación de la región con año $t-1$ y t , edad x y generaciones $g-1, g$ y $g+1$ del diagrama de Lexis.



Fuente: Elaboración propia

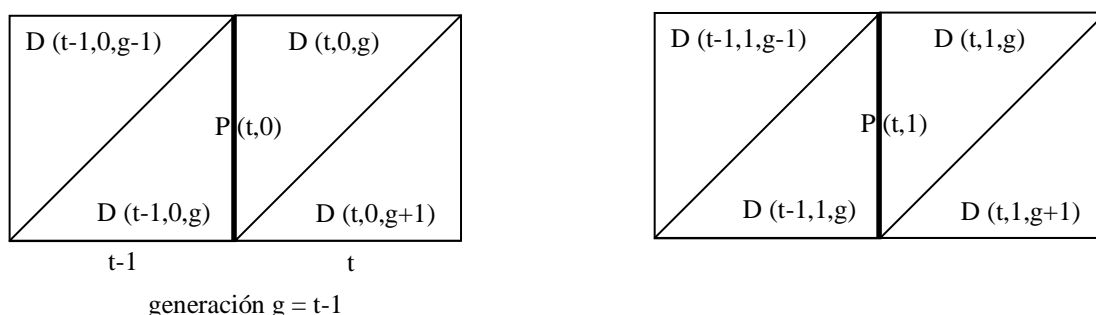
El riesgo de muerte para cada edad x se estima por la expresión:

$$q(x) = \frac{\frac{1}{2} \cdot (D(t,x) + D(t-1,x))}{P(t,x) + \frac{1}{2} \cdot D(t-1,x)} \quad (2.9)$$

Para el caso particular de las defunciones de 0 años, debido a que la mortalidad incide más en momentos próximos al nacimiento, se considera la hipótesis de que la proporción de fallecidos es superior en los triángulos inferiores de los

cuadrados. Es decir, $2/3$ de los fallecidos nacieron en el año considerado, $D(t-1,0,g)$ y $D(t,0,g+1)$ en la representación de la izquierda de la Figura 2.7, y el tercio restante en el año anterior, $D(t-1,0,g-1)$ y $D(t,0,g)$ en la representación de la izquierda de la Figura 2.7.

Figura 2.7. Representación de la región del año t-1 y t y edad x en el diagrama de Lexis.



Fuente: Elaboración propia

Para el caso particular de las defunciones de 1 años se considera la hipótesis de que $3/5$ de los fallecidos nacieron en el año anterior, $D(t-1,1,g)$ y $D(t,1,g+1)$ en la representación de la derecha de la Figura 2.7, y los $2/5$ restantes en el año precedente, $D(t-1,1,g-1)$ y $D(t,1,g)$ en la representación de la derecha de la Figura 2.7.

Las funciones de supervivientes y defunciones se calculan de acuerdo con las ecuaciones 2.1 y 2.2:

$$d(x) = q(x) \cdot l(x) \quad \text{y} \quad l(x+1) = l(x) - d(x) \quad (2.10)$$

Suponiendo que las defunciones son uniformes a lo largo del año la esperanza de vida se estima como:

$$e(x) = \frac{1}{2} + \frac{1}{l(x)} \cdot \sum_{k=x+1}^w l(k) \quad (2.11)$$

En 1951 se publican la serie de tablas relativas a 1900, 1910, 1920, 1930 y 1940 utilizando la misma metodología (INE, 1951). En 1960 se publica la tabla de mortalidad para 1950 (INE, 1960a), con metodología análoga a las anteriores

estimándose el riesgo de muerte a las edades de 0 y 1 de acuerdo con las expresiones:

$$q(0) = \frac{0,7 \cdot D(t-1,0) + 0,3 \cdot D(t,0)}{N(t-1)} \quad (2.12)$$

$$q(1) = \frac{0,6 \cdot D(t-1,1) + 0,4 \cdot D(t,1)}{N(t-2) - 0,7 \cdot D(t-2,0) - 0,3 \cdot D(t-1,1)} \quad (2.13)$$

donde $N(t)$ son los nacimientos en el año t .

Para las edades superiores a 75 años se interpoló una parábola de segundo grado entre las edades 70, 75 y 110 donde los valores del riesgo en las dos primeras edades son los obtenidos en la fórmula 2.9 y para la última se toma la unidad.

En 1977 se publican las tablas de mortalidad de 1960 y 1970 (INE, 1977a). En este caso los riesgos de muerte a los 0 y 1 año se estiman por:

$$q(0) = \frac{0,7 \cdot D(t-1,0) + 0,3 \cdot D(t,0)}{P(t,0) + 0,7 \cdot D(t-1,0)} \quad (2.14)$$

$$q(1) = \frac{0,6 \cdot D(t-1,1) + 0,4 \cdot D(t,1)}{P(t,1) - 0,6 \cdot D(t-1,1)} \quad (2.15)$$

Para edades superiores a 90 años se ajusta una parábola de tercer grado por mínimos cuadrados sobre las edades de 90 a 94 sujeto a las tres condiciones siguientes:

1. La curva pasa por $q(90)$ calculado de acuerdo con la fórmula 2.9.
2. La curva pasa por $q(110)$ que toma el valor unidad.
3. La curva tiene tangente paralela al eje x en la edad 110.

Estas restricciones hacen que en los datos publicados (INE, 1977a) a partir de la edad 90 los riesgos de muerte decrezcan en el primer tramo para luego volver a crecer.

En 1981 se publica la tabla de mortalidad para los años 1975-1976, (INE, 1981a). A partir de 1975, el MNP recoge las cifras de fallecidos clasificadas por edad y generación. Por otra parte, la población a 31 de diciembre de 1975 se toma de la

publicación “Evolución de la población de España. 1961-1978” (INE, 1979), donde se describe la estimación de los mismos. El punto de partida de la metodología es de nuevo el cálculo del riesgo de muerte. En este caso, como se puede observar en la Figura 2.6, a través de la expresión:

$$q(x) = \frac{D(t-1, x, g) + D(t, x, g)}{P(t, x) + D(t-1, x, g)} \quad (2.16)$$

Para la edad 0 se estima por:

$$q(0) = \frac{D(t-1, 0, g) + D(t, 0, g)}{N(t, 0)} \quad (2.17)$$

La serie de supervivientes y defunciones se calcula como siempre y en la metodología se incorpora la población estacionaria, $L(x)$, que se estima como:

$$L(x) = \frac{13}{24} \cdot (l(x) + l(x+1)) - \frac{1}{24} \cdot (l(x-1) + l(x+2)) \quad (2.18)$$

Para las edades de 1 a 88 años de acuerdo con el cálculo del área bajo la curva de supervivientes, $l(x)$, se aplica una fórmula de cuadratura descrita en Keyfitz (1979), con los siguientes casos particulares:

$$L(0) = \alpha(0) \cdot l(0) + \alpha(1) \cdot l(1) \quad (2.19)$$

$$\text{Donde } \alpha(0) = \frac{D(t, 0, g-1)}{D(t, 0, g-1) + D(t, 0, t)} \quad \text{siendo} \quad \alpha(1) = 1 - \alpha(0)$$

$$L(89) = e(89) \cdot l(89) - e(90) \cdot l(90) \quad \text{y}$$

$$L(90) = l(90) \cdot e(90)$$

La esperanza de vida en esta ocasión se calcula como:

$$e(x) = \frac{1}{l(x)} \cdot \sum_{k=x}^{90} L(k) \quad (2.20)$$

En el año 1988 se publica la tabla de mortalidad de la población española 1980-81. En este caso se vuelve a utilizar la metodología de 1977 para las tablas de 1960 y 1970 con la particularidad de que la población estacionaria se cierra a 100 años y más.

En 1991 se publica la tabla de mortalidad de la población española 1985-1986 (INE, 1991), cuya metodología no difiere de la anterior salvo en la estimación del riesgo de muerte a los 0 y 1 años para los que se utilizan las expresiones:

$$q(0) = \frac{\alpha(0) \cdot D(t-1,0) + \beta(0) \cdot D(t,0)}{P(t,0) + \alpha(0) \cdot D(t-1,0)} \quad (2.21)$$

$$q(1) = \frac{\alpha(1) \cdot D(t-1,1) + \beta(1) \cdot D(t,1)}{P(t,1) - \alpha(1) \cdot D(t-1,1)} \quad (2.22)$$

donde:

- $\alpha(0)$ es la proporción de fallecidos en el año t-1 con menos de un año de edad entre los nacidos ese año.
- $\beta(0)$ es la proporción de fallecidos en el año t con menos de un año de edad entre los nacidos el año t-1.
- $\alpha(1)$ es la proporción de fallecidos en el año t-1 con un año cumplido de edad entre los nacidos el año t-2.
- $\beta(1)$ es la proporción de fallecidos en el año t con un año cumplido de edad entre los nacidos el año t-2.

Se cumple además que $\alpha(0) + \beta(0) = 1$ y $\alpha(1) + \beta(1) = 1$.

En 1993 se publica la tabla de mortalidad de la población española del periodo 1990-91 (INE, 1993). La metodología empleada no difiere de la utilizada en las tablas referidas al año 1985.

En los años 1998, 1999 y 2002 se publican también, con la misma metodología, las tablas de mortalidad 1994-95, 1996-97, 1998-1999.

En julio de 2007 se publica la serie de tablas de mortalidad desde 1992 hasta 2005 (INE, 2007), siguiendo la metodología análoga a la anterior, salvo que el riesgo de muerte para las edades 0 y 1 se estima de forma diferente según las expresiones:

$$q(0) = \frac{D(t-1,0, g) + D(t,0, g)}{P(t,0) + D(t-1,0, g)}$$

$$q(1) = \frac{D(t-1,1, g-1) + D(t,1, g-1)}{P(t,1) + D(t-1,1, g-1)}$$

para la que se cumple que $g=t-1$.

Con estas tablas concluye una serie que sigue el mismo principio desde 1945 de calcular el riesgo de muerte en base a la observación de la mortalidad durante dos años consecutivos.

La metodología utilizada por el INE hasta esa fecha se plantea la tabla de mortalidad como un objetivo partiendo de los datos de defunciones y sobre todo de población a partir de los censos disponibles, referenciando la tabla a una fecha concreta de 1 de enero del año $t+1$, y tomando datos observados de los dos años t y $t+1$.

Ante unas circunstancias que han cambiado, es necesario considerar un nuevo enfoque que tenga en cuenta la creciente relevancia de la estimación de la población a 1 de enero de cada año t . Para proyectar el stock de efectivos a 1 de enero del año t al 1 de enero del año $t+1$ es necesario utilizar un indicador que mida la intensidad de la mortalidad durante todo el año t .

El 21 de septiembre de 2009 se celebró en Madrid un Seminario sobre tablas de mortalidad en el que participaron expertos nacionales y extranjeros con el objetivo de desarrollar nuevos avances metodológicos en el cálculo de las tablas de mortalidad en España. Aprovechando las conclusiones del seminario, en 2010 el INE publicó la serie de tablas para España desde 1981 de acuerdo con una nueva metodología, en

vigor hasta comienzos de 2013⁷, y que parte de la observación únicamente del año objeto de estudio. Esta metodología se inspiraba en los principios, conceptos y definiciones del protocolo metodológico de Human Mortality Database, que se presenta a continuación.

2.5. Protocolo metodológico de Human Mortality Database

El protocolo metodológico elaborado por Human Mortality Database (en adelante HMD) supone un importante esfuerzo de integración para dar respuesta al vacío de un marco común en la medición de la intensidad de la mortalidad.

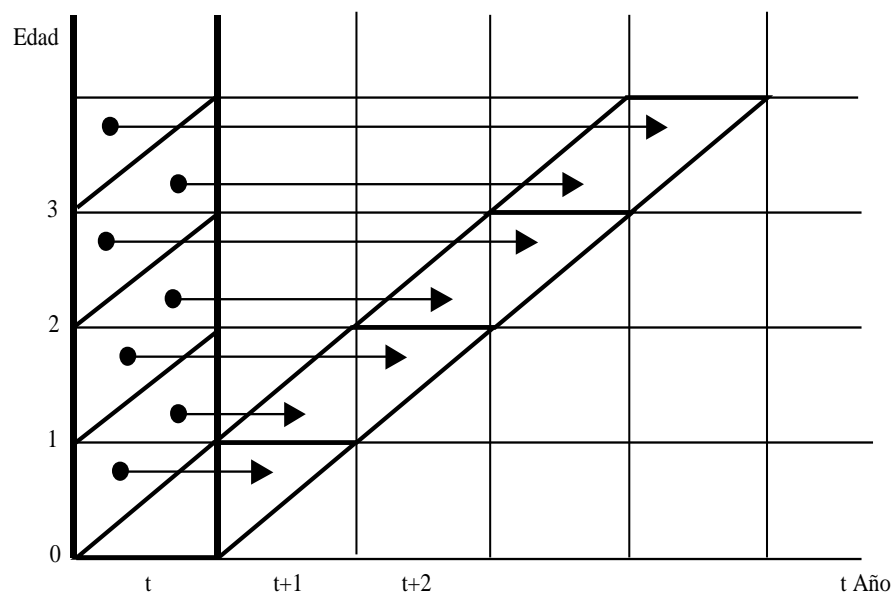
Este proyecto de colaboración se puso en marcha en el año 2000 y en él participan equipos de investigación de reconocido prestigio pertenecientes al Departamento de Demografía de la Universidad de Berkeley, Estados Unidos, y al Instituto Max Planck de Investigación Demográfica en Rostock, Alemania. El proyecto desarrolla un protocolo metodológico común, coherente y sistemático para facilitar la comparación de la incidencia de la mortalidad entre los distintos países. Dicho protocolo es resultado de la colaboración, discusión y debate de los principales expertos en la materia. Adicionalmente, su objetivo final es elaborar una base de datos a nivel mundial con la información necesaria para construir las tablas de mortalidad de cualquier país.

La metodología de la HMD⁸ se basa en el estudio de la mortalidad de momento para un año t determinado. No existe una fecha de referencia sino un periodo de un año a diferencia de las tablas de mortalidad que se han elaborado históricamente en España. En la Figura 2.8 se observa cómo se construye la generación ficticia representada en el diagrama de Lexis a partir de los triángulos de la columna determinada por el año t y debidamente trasladados.

⁷ En junio de 2013 se publicó la serie de tablas actualizada en el periodo 1991-2011. Estas tablas utilizan la metodología (2013) para el periodo 2002-2011 y el resto sigue con la metodología (2010).

⁸ Véase Wilmoth, Andreev, Jdanov y Glej (2007).

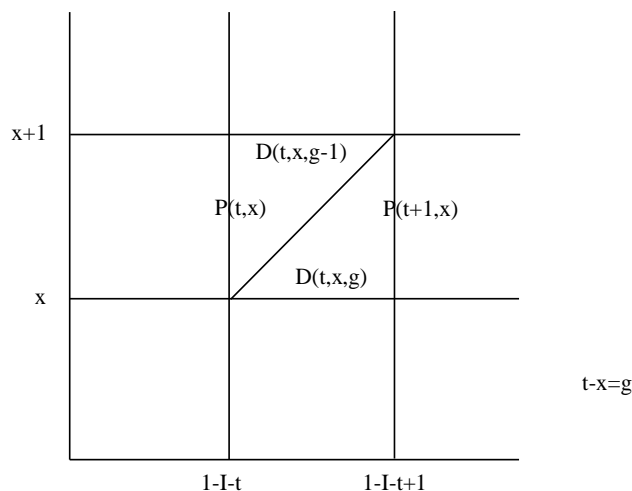
Figura 2.8. Diagrama de Lexis representando la traslación de la columna a la generación ficticia.



Fuente: Elaboración propia

Bajo los criterios establecidos por el protocolo metodológico de HMD, no se pueden calcular directamente probabilidades de muerte, ya que esta función está asociada en el diagrama de Lexis a paralelogramos y no a los cuadrados como el que se observa en la Figura 2.9, cuando el objeto de estudio es el periodo. En este caso, primero se calcula la tasa de mortalidad $m(x)$, que es la medida asociada a los cuadrados del diagrama de Lexis. A partir de la intensidad descrita por las tasas, se considera una generación ficticia que es alcanzada por esa intensidad para la que se calculan el resto de las funciones de la tabla de mortalidad.

Figura 2.9. Diagrama de Lexis. $P(t, x)$ es el total de líneas de vida que cortan la isócrona a 1 de enero del año t , o población a $1-I-t$ con x años. $D(t, x, g)$ es el total de líneas de vida que terminan en el triángulo de año t , edad x y generación g o total de defunciones el año t con x años de la generación g .



Fuente: Elaboración propia

Por tanto, el punto de partida del cálculo de una tabla de mortalidad es la obtención de la tasa de mortalidad de la población real en el año t .

Se define la tasa de mortalidad $m(t, x, s)$ para el año t a la edad x y el sexo s , como un cociente donde el numerador es el número de defunciones ocurridas en el año t de individuos con edad x y sexo s , $D(t, x, s)$ y el denominador es la exposición al riesgo, $E(t, x, s)$, o total de tiempo, en años, vividos por los individuos con edad x en el año t de sexo s , es decir, años-persona de los individuos de edad x durante el año t de sexo s y se tiene que:

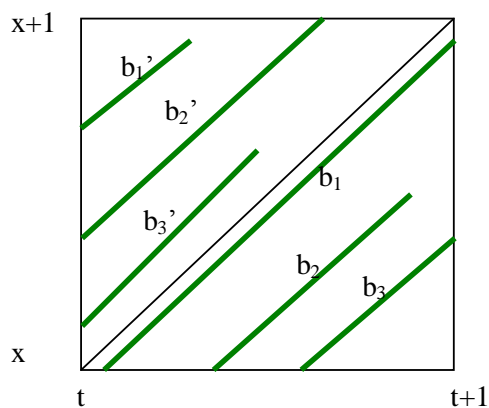
$$m(t, x, s) = \frac{D(t, x, s)}{E(t, x, s)} \quad (2.23)$$

El numerador de $m(t, x, s)$ se calcula como:

$$D(t, x, s) = D(t, x, g - 1, s) + D(t, x, g, s) \quad (2.24)$$

donde $D(t, x, g, s)$ es el número de individuos fallecidos en el año t , con edad x , de la generación g y sexo s y coincide con el total de las líneas de vida que terminan en el triángulo superior de la Figura 2.10.

Figura 2.10. Región del diagrama del Lexis determinada por la edad x y el año t .



Fuente: Elaboración propia

En la Figura 2.10 el triángulo inferior corresponde a la generación $g=t-x$ y el triángulo superior a la generación $g-1$. Los segmentos en trazo grueso verde son la parte de las líneas de vida de distintos individuos con edad x en el año t . Los individuos representados por b_1' , b_3' y b_2 fallecen a la edad x durante el año t . Sin embargo, los individuos b_1 , b_3 y b_2' sobreviven a la edad x o al año t .

El denominador de $m(t, x, s)$ o exposición al riesgo, $E(t, x, s)$, se calcula como la suma de todos los tiempos de exposición al riesgo de la generación $g=t-x$ y la generación $g-1$:

$$E(t, x, s) = E(t, x, g - 1, s) + E(t, x, g, s) \quad (2.25)$$

El tiempo de la línea de vida se mide en tiempo cronológico, es decir, no en la longitud del segmento sino en la distancia definida como valor absoluto de la diferencia en las abscisas y que es equivalente al valor absoluto de la diferencia en las ordenadas:

$$d(\underline{x}, \underline{y}) = |y_2 - x_2| = |y_1 - x_1| \text{ siendo los puntos } \underline{x} = (x_1, x_2), \underline{y} = (y_1, y_2)$$

La definición de la expresión 2.23 es válida para poblaciones abiertas expuestas a las migraciones. Cualquier inmigrante que llega a la población objeto de estudio en el año t se registra y su línea de vida se incluye en el diagrama de Lexis en su correspondiente cuadrado en función de la edad x . Así mismo, los emigrantes que se marchan de la población dejan de contar en la exposición al riesgo de la población. Esta es una visión dinámica de la población más acorde con la realidad.

El segundo sumando de la fórmula 2.23 correspondiente al triángulo inferior del diagrama de Lexis, y se calcula como:

$$E(t, x, g, s) = \sum_i^{P(t+1, x, s) + D(t, x, g, s)} b(t, x, g, s, i) \quad (2.26)$$

donde $b(t, x, g, s, i)$ es el tiempo de vida vivido por el individuo i durante el año t cuando tenía x años. Para cada individuo i de los $D(t, x, g, s)$ fallecidos con edad x el año t de la generación g se define como el tiempo en años desde que cumple x años hasta que fallece. Para cada individuo i de los $P(t+1, x, s)$ supervivientes a 1 de enero del año $t+1$ es ésta fecha menos la fecha de su x cumpleaños.

El primer sumando de la fórmula 2.25 correspondiente al triángulo superior de la Figura 2.10 se calcula como:

$$E(t, x, g - 1, s) = \sum_i^{P(t, x, s)} b'(t, x, g - 1, s, i) \quad (2.27)$$

donde $b'(t, x, g - 1, s, i)$ es el tiempo de vida vivido por el individuo i de la generación $g-1$ durante el año t cuando tenía x años. Para cada individuo i de los $D(t, x, g - 1, s)$ fallecidos con edad x el año t de la generación $g-1$ se define como el tiempo en años desde el 1 de enero del año t hasta que fallece. Para cada individuo i de los $P(t, x, s) - D(t, x, g - 1, s)$ supervivientes que llegan a cumplir los $x+1$ es la fecha del $x+1$ cumpleaños menos el 1 de enero del año t .

A cada edad y para cada sexo, las tasas de mortalidad realmente observadas en la población tienen que ser iguales a las de la población estacionaria subyacente en la generación ficticia.

$$m(x) = \frac{D(x)}{E(x)} = \frac{d(x)}{L(x)}. \quad (2.28)$$

A partir de la tasa se calcula el riesgo $q(x)$, se entiende que se refiere al año t y el sexo s , como:

$$q(x) = \frac{m(x)}{1 + (1 - a(x)) \cdot m(x)} \quad (2.29)$$

donde la expresión se deduce de la forma siguiente:

Sea $q(x) = \frac{d(x)}{l(x)}$ y $m(x) = \frac{d(x)}{L(x)}$; a partir de la fórmula 2.3 y teniendo en

cuenta que $l(x+1) = l(x) - d(x)$:

$$L(x) = l(x) - d(x) + a(x) \cdot d(x) \quad (2.30)$$

se tiene que:

$$\frac{q(x)}{m(x)} = \frac{l(x) - d(x) + a(x) \cdot d(x)}{l(x)} = 1 - (1 - a(x)) \cdot q(x). \quad (2.31)$$

Entonces:

$$q(x) = m(x)[1 - (1 - a(x)) \cdot q(x)] = m(x) - (1 - a(x)) \cdot q(x) \cdot m(x)$$

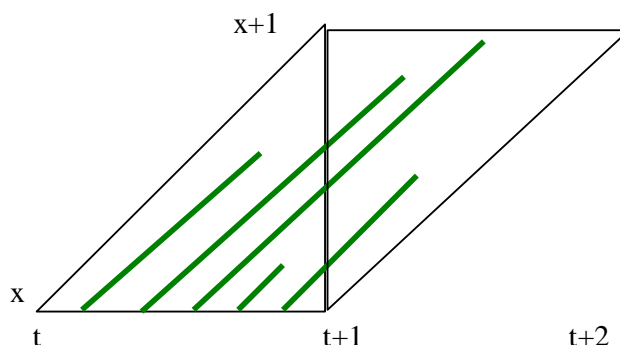
$$q(x) + (1 - a(x)) \cdot q(x) \cdot m(x) = m(x)$$

$$q(x)[1 + (1 - a(x)) \cdot m(x)] = m(x)$$

Y despejando se tiene la fórmula 2.29.

El tiempo vivido el último año de vida, $a(x)$, promedio del tiempo vivido por todos los individuos que fallecen, solo puede calcularse en estricto sentido al observar longitudinalmente como se aprecia en la Figura 2.11.

Figura 2.11. Región del diagrama de Lexis determinada por la edad x y la generación $t-x$. $a(x)$ es el promedio de todos los segmentos de trazo grueso verde.

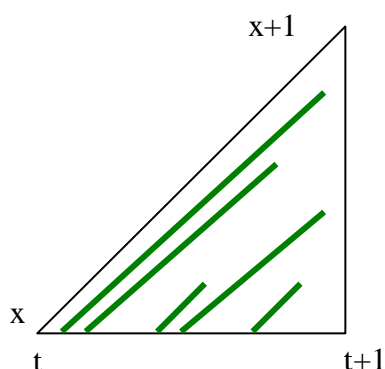


Fuente: Elaboración propia

Para calcular $a(x)$ en una región de la forma de la Figura 2.10 hay que observar los fallecidos en el triángulo inferior de la generación $g=t-x$ y los fallecidos en el triángulo superior de la generación $g-1$.

En el triángulo inferior es inmediato que coincide con los segmentos $b(t, x, g, i)$.

Figura 2.12. Segmentos $b(t, x, g, i)$, en trazo grueso verde, en la región del diagrama de Lexis determinada por la edad x y el año t y la generación $g=t-x$.

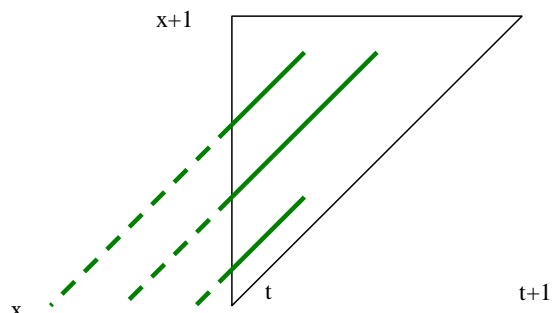


Fuente: Elaboración propia

En el triángulo superior de la Figura 2.10 para cada individuo i el segmento $b'(t, x, g-1, i)$ se extiende desde la fecha de cumpleaños, como se observa en la Figura 2.12, para estimar el tiempo vivido el último año de vida del individuo i $a(t, x, g-1, i)$. Además se tiene que:

$$a(t, x, g - l, i) = l + b(t, x, g - l, i) \quad (2.32)$$

Figura 2.13. Extensión de los segmentos b'.



Fuente: Elaboración propia

Una vez calculada la $q(x)$ y $a(x)$ las demás funciones de la tabla de mortalidad, $l(x)$, $d(x)$, $L(x)$, $T(x)$ y $e(x)$ son inmediatas de obtener, como ya se ha comentado previamente.

El cálculo de una tabla de mortalidad a partir de la estimación con la fórmula 2.23 requiere de un registro con información exhaustiva para cada individuo de la población objeto de estudio.

Sin embargo, estos requisitos son demasiado exigentes para la gran mayoría de países porque requieren de un censo continuo a lo largo del tiempo, que es algo que no se dispone en los países en general. Solo países nórdicos como Suecia o Finlandia disponen de algo parecido con registros suficientemente potentes de información tan precisa y detallada para todos sus habitantes.

Para que la base de datos abarque el máximo posible de países, HMD pide unos requisitos menos exigentes a las fuentes de información. Se debe disponer de datos de población desagregado por edad y sexo a 1 de enero de cada año y defunciones por año, generación, edad simple y sexo.

Con estas restricciones, el protocolo metodológico del HMD parte de las siguientes hipótesis:

1. Las líneas de vida de los individuos de las generaciones g y $g-1$ cortan uniformemente las isócronas a 1 de enero del año $t+1$ y del año t respectivamente⁹.
2. Las defunciones se distribuyen uniformemente respecto al tiempo, la edad y la generación, i.e. dentro de cada triángulo la distribución es uniforme¹⁰.
3. La población objeto de estudio esta cerrada a las migraciones.

Las hipótesis del HMD son una simplificación de la realidad que responden a las limitaciones de las fuentes de información disponibles en muchos países.

Partiendo de estas hipótesis, se calculan las funciones biométricas de la tabla de mortalidad comenzando por una estimación de la tasa de mortalidad.

Se estima la tasa de mortalidad $m(t, x, s)$ para el año t a la edad x y el sexo s , por la fórmula 2.21 donde la exposición al riesgo se estima mediante:

$$E(t, x, s) = \frac{1}{2} \cdot P(t, x, s) + \frac{1}{2} \cdot P(t+1, x, s) + \frac{1}{6} (D(t, x, g, s) - D(t, x, g-1, s)) \quad (2.33)$$

Esta expresión se deduce del siguiente razonamiento:

El tiempo que vive en media cada individuo de los $P(t, x, s) - D(t, x, g-1, s)$ que, teniendo x años a 1 de enero del año t , llega a su $x+1$ cumpleaños es $\frac{1}{2}$ años.

Con lo que la exposición al riesgo de estos individuos es $\frac{1}{2} \cdot [P(t, x, s) - D(t, x, g-1, s)]$ años-persona.

⁹ Con información suficiente sobre la fecha de los nacimientos de cada generación, etc. se pueden contrastar otras hipótesis sobre la verdadera distribución. Una vía abierta de investigación es determinar la distribución más aproximada utilizando como muestra la fecha de nacimiento de la estadística de nacimientos y la estadística de defunciones en España.

¹⁰ Análogamente, en el caso de no disponer de la verdadera distribución de las defunciones, con información sobre la fecha de las defunciones de cada generación se pueden contrastar otras distribuciones alternativas a la uniforme, como por ejemplo una distribución beta.

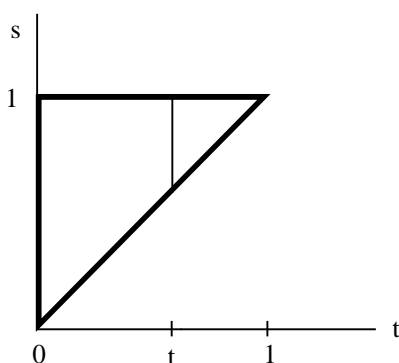
A esta exposición hay que sumarle el tiempo que viven los individuos que fallecen en el triángulo superior. Sea T la variable aleatoria que representa el tiempo cronológico transcurrido el último año y S la variable aleatoria que representa el tiempo vivido el último año de vida en el triángulo del diagrama de Lexis de la Figura 2.14 con distribución uniforme en la región. La esperanza de esta variable es $E(T, S) = (1/3, 2/3)$.

$$E[T] = \iint_U t \cdot f(t, s) \cdot ds dt = \iint_U t \cdot 2 \cdot ds dt = \int_0^1 ds \int_0^s 2t dt = \int_0^1 ds \cdot t^2 \Big|_{t=0}^{t=s} = \int_0^1 s^2 ds = \frac{s^3}{3} \Big|_{s=0}^{s=1} = 1/3$$

$$E[S] = \iint_U s \cdot f(t, s) ds dt = \iint_U 2 \cdot s ds dt = \int_0^1 ds \int_0^s 2s dt = \int_0^1 2s^2 ds = \frac{2}{3} \cdot s^3 \Big|_{s=0}^{s=1} = 2/3$$

De esto se deduce que el tiempo medio vivido en la región es $1/3$.

Figura 2.14. Región U del plano limitada por el eje de ordenadas, la bisectriz del primer cuadrante y la recta $s = 1$.



Fuente: Elaboración propia

El total de años-persona vividos por los fallecidos por la generación $g-1$ es

$$\frac{1}{3} \cdot D(t, x, g - 1, s)$$

Luego la exposición al riesgo en el triángulo superior es:

$$E(t, x, g - 1, s) = \frac{1}{2} \cdot [P(t, x, s) - D(t, x, g - 1, s)] + \frac{1}{3} \cdot D(t, x, g - 1, s) = \frac{1}{2} \cdot P(t, x, s) - \frac{1}{6} \cdot D(t, x, g - 1, s) \quad (2.34)$$

Para el triángulo inferior se tiene que:

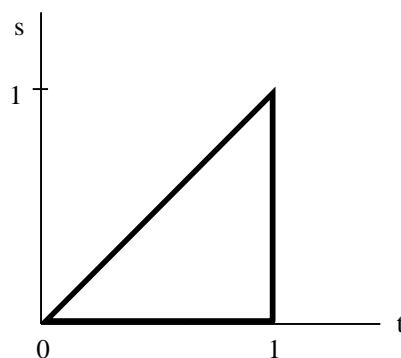
El tiempo que vive en media cada individuo de los $P(t+I, x, s)$ que llegan al 1 de enero del año $t+1$ es $\frac{I}{2}$ años. Con lo que la exposición al riesgo de estos individuos es $\frac{I}{2} \cdot P(t+I, x, s)$ años-persona.

A este tiempo hay que añadirle el tiempo que viven los individuos que fallecen en el triángulo inferior de la Figura 2.15. En este caso $E(T, S) = (2/3, 1/3)$.

$$E[T] = \iint_L t \cdot f(t, s) \cdot ds dt = \iint_L t \cdot 2 \cdot ds dt = \int_0^1 dt \int_0^t 2t ds = \int_0^1 dt \cdot 2s^2 \Big|_{s=0}^{s=t} = \int_0^1 2t^2 dt = 2 \cdot \frac{s^3}{3} \Big|_{s=0}^{s=1} = 2/3$$

$$E[S] = \iint_L s \cdot f(t, s) ds dt = \iint_L 2 \cdot s ds dt = \int_0^1 dt \int_0^t 2s ds = \int_0^1 t^2 dt = \frac{1}{3} \cdot t^3 \Big|_{t=0}^{t=1} = 1/3$$

Figura 2.15. Región L del plano limitada por el eje de abscisas, la bisectriz del primer cuadrante y la recta $t = 1$.



Fuente: Elaboración propia

Para un individuo que muere a los $1/3$ años desde el último cumpleaños cuando han transcurrido $2/3$ años de ese año cronológico el tiempo vivido en la región L es $2/3 - 1/2 = 1/6$:

El total de años-persona vividos por los fallecidos por la generación g es

$$\frac{I}{6} \cdot D(t, x, g, s)$$

Luego la exposición al riesgo en el triángulo superior es:

$$E(t, x, g, s) = \frac{I}{2} \cdot P(t+I, x, s) + \frac{I}{6} \cdot D(t, x, g, s) \quad (2.35)$$

Sumando la exposición al riesgo de los dos triángulos se obtiene la expresión 2.33.

Una vez calculada la tasa de mortalidad en el año t a cada edad x y sexo s $m(t, x, s)$ se estima la probabilidad o riesgo de muerte el año t con edad x y sexo s , $q(t, x, s)$, de una generación ficticia que presenta la misma incidencia de la mortalidad a cada edad y en cada sexo que la población observada en el año de estudio t por la expresión:

$$q(t, x, s) = \frac{m(t, x, s)}{1 + (1 - a(t, x, s)) \cdot m(t, x, s)} \quad (2.36)$$

donde $a(t, x, s)$ es el tiempo promedio vivido el año t con edad x por los individuos de sexo s de la población en estudio que mueren con dicha edad. Se calcula como media ponderada de la duración esperada de fallecidos en cada triángulo de los x años, por la expresión:

$$a(t, x, s) = \frac{1/3 \cdot D(t, x, g, s) + 2/3 \cdot D(t, x, g - 1, s)}{D(t, x, g, s) + D(t, x, g - 1, s)} \quad (2.37)$$

A partir del riesgo de muerte se calcula el resto de funciones de la tabla de mortalidad. El riesgo de muerte y cada una de estas funciones definidas para cada sexo sobre dos dimensiones, tiempo cronológico y edad, denominadas superficies que se dicen de mortalidad.

2.6. Protocolo metodológico actual de la medición de la incidencia de la mortalidad en España

Actualmente en España no se dispone de un registro de población exhaustivo a modo de censo continuo para el cálculo exacto de las tablas de mortalidad, como es el caso de Suecia. Lo más parecido es el padrón municipal, un registro permanente que permite el seguimiento constante de cada individuo de la población, lo que ahorraría todo el trabajo de estimación y proyección. El padrón es una herramienta

muy valiosa como primera aproximación al estudio de la población, de hecho proporciona el dato oficial, pero tiene sus limitaciones. Por ejemplo, los extranjeros extracomunitarios que llegan a España suelen empadronarse porque conlleva una serie de derechos, como asistencia sanitaria o educativa de los hijos, pero cuando se van no se dan de baja. El último censo descubre que el padrón contabiliza personas que en realidad ya no residen en España. El censo de 2011 da una población de 46.818.216 residentes a 1 de enero de 2012¹¹ y la sobreestimación del padrón se va hasta los 47.265.321 residentes, 447.105 personas más. Hay distintos tipos de errores en el padrón, desde el momento en que no se registran todas las personas que llegan o los propios españoles no cambian siempre el lugar de residencia aunque se muden. Estas inconsistencias llevan a la necesidad de estimar el tamaño real de la población.

En España el organismo que se encarga de construir las tablas de mortalidad es el Instituto Nacional de Estadística y las elabora para ambos sexos a nivel nacional, por provincias y por Comunidades Autónomas.

La publicación de las tablas de mortalidad para la serie 2002-2013 incorpora la última actualización de la metodología INE (2013). El resto de la serie desde 1991 calcula las tablas de mortalidad siguiendo la metodología INE (2009) a partir de 2010. Una metodología sujeta a los principios del protocolo metodológico de la Base de Datos de Mortalidad Humana (Human Mortality Database), adaptados a la demanda de los usuarios y las limitaciones de las fuentes de información disponibles en España.

Puestos a pedir, los usuarios quieren tablas de mortalidad a cualquier nivel de desagregación territorial, ciudad e incluso barrio. Ocurre que la estadística en general, y las estimaciones de las tablas de mortalidad en particular, son una ciencia para grandes números. Tiene poco sentido calcular la tabla de mortalidad de una población de pequeño tamaño. Sea $\{X(t,x)\}$ el proceso estocástico de variables aleatorias definidas como el número de muertos de edad x el año t en una población. $X(t,x)$ sigue una distribución de Poisson de parámetro $m(t,x) \cdot E(t,x)$, producto de la

¹¹ A partir de los datos del censo a 1 de noviembre de 2011: 46.815.916 residentes. En la web: www.ine.es

tasa de mortalidad por la exposición al riesgo (Camarda, 2008). La probabilidad de cero defunciones de edad x es:

$$P[X = 0] = \exp\{-m(t, x) \cdot E(t, x)\}$$

Esto quiere decir que es factible que ocurra el suceso. Cuando la población es grande, $E(t, x)$ es grande y la probabilidad de cero defunciones tiende a cero. Pero cuanto más pequeño sea $m(t, x)$, mayor es la probabilidad de cero defunciones. Es difícil que se observen ceros en el riesgo de muerte en tablas de mortalidad para poblaciones grandes. A nivel nacional, con un tamaño de la población, tanto para varones como para mujeres, por encima de los 20 millones, no se han observado ceros en los 18 años analizados desde 1991 hasta 2008. En realidad no es tan descabellado. Puede ocurrir en un futuro relativamente próximo. De hecho los datos así lo auguran con menos de 20 defunciones de niñas en algunas edades y disminuyendo.

La existencia de riesgos nulos observados, no necesariamente repetida en la misma edad, ya se da en todas las provincias incluso en la más grande como Madrid. Con un tamaño poblacional superior a los 3 millones de varones se observan cero defunciones el año 2006 con 11 años y el año 2009 con 7 años. El fenómeno crece al disminuir el tamaño de la población alcanzando los 332 ceros observados para distintas edades en riesgo de muerte para los varones de Melilla con un tamaño poblacional inferior a 35.000 individuos.

Multitud de factores circunstanciales o casualidades alteran la estimación de los parámetros sobremanera. En una población en la que en un año se mueren uno y al año siguiente se mueren dos individuos a una determinada edad, como es fácil que suceda en poblaciones pequeñas, el riesgo de muerte se duplica de un año a otro. O incluso se hace cero si se tiene la suerte circunstancial de que no haya ninguna defunción a esa edad.

De acuerdo con la división político-administrativa de España es obligatorio el cálculo de las tablas de mortalidad a nivel provincial. La burocracia no para a pensar

que la estimación en provincias con población de pequeño tamaño de problemas y sufra por falta de consistencia.

La fuente de información para los fallecidos es la estadística de defunciones del registro del Movimiento Natural de Población (MNP). El MNP es una estadística de larga tradición en España. La primera publicación data de 1858 por la Junta General de Estadística del Reino, antecesor del INE. Tiene carácter anual y se elabora a partir de la información suministrada mensualmente por los Registros Civiles.

La estadística proporciona información del día, mes, año de nacimiento y defunción, y la provincia de residencia de todos los residentes en España fallecidos, tanto españoles como extranjeros. Cada defunción se anota en el Registro Civil, y los Ayuntamientos se encargan de recopilar y facilitar mensualmente a la Delegación Provincial del INE el boletín estadístico de defunciones, junto con el de nacimientos y matrimonios.

Los Servicios Centrales del INE reciben los boletines a lo largo del año en curso t . La Subdirección General de Estadísticas de Población se encarga de reunir, procesar y depurar los registros durante el año siguiente, $t+1$. Los datos del MNP se dan por definitivos a finales del año $t+1$ o comienzos del año posterior, $t+2$, en el que se utilizan para elaborar la tabla de mortalidad observada del año t . Para los stocks de población referidos a 1 de enero de cada año el INE calcula las Estimaciones de Población Actual (ePOBa), estimación que se hace a tiempo real.

Todos los años terminados en 1 se realiza, en España, un censo de población. Un censo es una observación exhaustiva de toda la población que se toma como el verdadero valor del stock de población.

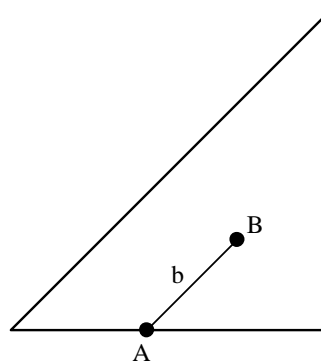
A partir de este dato se calcula el tamaño de la población a 1 de enero de cada año recurrentemente utilizando un procedimiento basado en el método de las componentes. Se parte de la población de tamaño conocido a comienzos de un año, se estima la evolución de los fenómenos demográficos básicos para el año en curso: (nacimientos, defunciones y migraciones) y se obtiene la población a comienzos del año siguiente.

La estimación de población actual proporciona el stock de la población por sexo y edad a 1 de enero de cada año con el grupo abierto a partir de 100 y más años para cada provincia¹².

La metodología del INE (2009) parte del supuesto de que la población objeto de estudio es una población cerrada, sin interferencias de las migraciones.

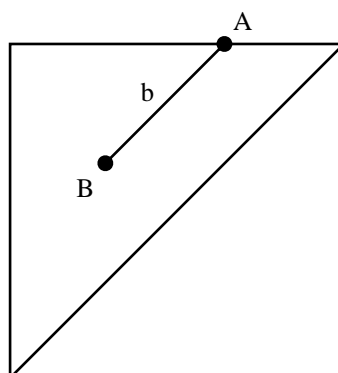
De acuerdo con la metodología INE (2009), para elaborar la tabla de mortalidad, en primer lugar hay que estimar la tasa de mortalidad por edad x , $m(t, x, s)$.

Figura 2.16. Triángulo inferior del diagrama de Lexis.



Fuente: Elaboración propia

Figura 2.17. Triángulo superior del diagrama de Lexis.



Fuente: Elaboración propia

¹² Las tablas de mortalidad a nivel provincial se cierran a 95 años excepto Ceuta y Melilla con cierre a 90.

El valor del segmento b se mide como valor absoluto de la diferencia de las abscisas entre B (fecha de defunción) y A (fecha de cumpleaños). En el triángulo inferior es positivo y en el triángulo superior es negativo.

La exposición al riesgo de la tasa de mortalidad bajo la hipótesis de distribución uniforme de corte de las líneas de vida de los individuos con las isócronas a 1 de enero del año t y $t+1$ resulta de la expresión:

$$E(t, x, s) = I/2 \cdot P(t, x, s) + I/2 \cdot P(t+1, x, s) + \sum_i b(t, x, s, i) \quad (2.38)$$

La fórmula 2.38 se deduce de la siguiente manera:

El tiempo vivido por los individuos en el cuadrado correspondiente al año t y edad x esta formado por los dos triángulos que se observan en la Figura 2.6, el inferior de la generación g y el superior de la generación $g-1$.

Los individuos de la generación g que llegan al 1 de enero del año $t+1$ viven en media medio año y su tiempo vivido es $I/2 \cdot P(t+1, x, s)$.

La exposición al riesgo de la generación g se calcula agregando a este tiempo, el tiempo vivido por los que se mueren y queda:

$$E(t, x, g, s) = I/2 \cdot P(t+1, x, s) + \sum_i b(t, x, g, s, i) \quad (2.39)$$

donde $b(t, x, g, s, i)$ es el tiempo vivido en el año t con edad x perteneciente a la generación g de sexo s por el individuo i .

En la Figura 2.16 se puede observar que dicha cantidad es positiva y coincide con $a(t, x, s, i)$, tiempo vivido (en años) en el año t con edad cumplida x de sexo s para cada individuo i , en el diagrama de Lexis incluido en la región delimitada por el triángulo inferior.

Además supone distribución uniforme de las líneas de vida de la generación $g-1$ a fecha 1 de enero del año t , se cumple que la suma de los segmentos b' , desde el 1 de enero del año t hasta la defunción, y b , desde la defunción hasta la fecha de cumpleaños, que no celebra, el año t , suman en media $I/2$, i.e.:

$$\sum_i (b'(t, x, g - 1, s, i) - b(t, x, g - 1, s, i)) = 1/2 \cdot D(t, x, g - 1, s) \quad (2.40)$$

Luego despejando:

$$-1/2 \cdot D(t, x, g - 1, s) + \sum_i b'(t, x, g - 1, s, i) = \sum_i b(t, x, g - 1, s, i) \quad (2.41)$$

La exposición al riesgo de la generación g-1 se calcula como:

$$\begin{aligned} E(t, x, g, s) &= 1/2 \cdot [P(t, x, s) - D(t, x, g - 1, s)] + \sum_i b'(t, x, g - 1, s, i) = \\ &= \frac{1}{2} \cdot P(t, x, s) - \frac{1}{2} \cdot D(t, x, g - 1, s) + \sum_i b'(t, x, g - 1, s, i) = \frac{1}{2} \cdot P(t, x, s) + \sum_i b(t, x, g - 1, s, i) \end{aligned}$$

Con lo que la exposición al riesgo de los $P(t, x, s)$ individuos de la generación g-1 que sobreviven a 1 de enero del año t se estima por $1/2 \cdot P(t, x, s) - \sum_i |b(t, x, g - 1, s, i)|$. En la Figura 2.15 para el triángulo superior se puede observar que dichos valores b son negativos. Operando el total de años vividos queda:

$$\begin{aligned} &1/2 \cdot P(t, x, s) - \sum_i |b(t, x, g - 1, s, i)| + 1/2 \cdot P(t + 1, x, s) + \sum_i b(t, x, g, s, i) = \\ &= 1/2 \cdot P(t, x, s) + 1/2 \cdot P(t + 1, x, s) + \sum_i b(t, x, g, s, i) - \sum_i |b(t, x, g - 1, s, i)| = \\ &= 1/2 \cdot P(t, x, s) + 1/2 \cdot P(t + 1, x, s) + \sum_i b(t, x, g, s, i) - [\sum_i -b(t, x, g - 1, s, i)] = \\ &= 1/2 \cdot P(t, x, s) + 1/2 \cdot P(t + 1, x, s) + \sum_i b(t, x, g, s, i) + \sum_i b(t, x, g - 1, s, i) = \\ &= 1/2 \cdot P(t, x, s) + 1/2 \cdot P(t + 1, x, s) + \sum_i b(t, x, s, i) \end{aligned}$$

De acuerdo con la metodología del INE (2009) el cierre de la tabla de mortalidad se hace a la edad 100.

Además se define la tasa de mortalidad en el año de estudio t para las edades de w y más años y sexo s como:

$$m(t, w+, s) = \frac{D(t, w+, s)}{1/2 \cdot P(t, w+, s) + 1/2 \cdot P(t + 1, w+, s) + \sum_{i, x \geq w} b(t, w+, s, i)} \quad (2.42)$$

Se estima la probabilidad o riesgo de muerte el año t con edad x y sexo s $q(t, x, s)$ de una generación ficticia que presenta la misma incidencia de la mortalidad a cada edad y en cada sexo que la población observada en el año de estudio t , por la expresión (2.4).

Con $a(t, x, s)$ en este caso estimada por la expresión:

$$a(t, x, s) = \frac{\sum_{i=1}^{D(t,x,s)} a(t, x, s, i)}{D(t, x, s)} \quad \text{para } x = 0, \dots, w-1. \quad (2.43)$$

Se establece por convenio que para el intervalo abierto final:

$$a(t, w+, s) = \frac{I}{m(t, w+, s)} \quad (2.44)$$

De esta manera, el riesgo de muerte en el intervalo abierto es uno.

A partir de la función $q(t, x, s)$, estimada para los valores de 0 a $w+$, se calcula el resto de funciones de la tabla de mortalidad con el protocolo metodológico de la Base de Datos de Mortalidad Humana.

2.7 Conclusiones y Alternativas Metodológicas

A lo largo de este capítulo, se ha revisado exhaustivamente el fenómeno demográfico de la mortalidad. El estudio se centra en la medición de su incidencia, utilizando la tabla de mortalidad como instrumento más adecuado. En el apartado 2.3 se recorre la evolución histórica de las metodologías de cálculo aplicadas en cada momento, restringidas a las limitaciones de las fuentes de información disponibles. Se culmina con el procedimiento teórico óptimo de cálculo de una tabla de mortalidad en condiciones ideales y bajo la filosofía imperante en la actualidad marcada por la metodología HMD. Una metodología que marca tendencia como denominador común de mínimos de cualquier metodología. Posteriormente en el apartado 2.6 se analiza el caso particular de la metodología del INE (2009), lo que da pie a describir

posibles mejoras, algunas incluidas en la metodología INE (2013) aplicada a la serie 2002-2011 que tratamos en este apartado.

Es importante hacer notar que cualquiera de las metodologías de estimación de tablas de mortalidad expuestas, incluidas las más consolidadas y generalmente aceptadas como el protocolo metodológico de HMD, aceptan que la estimación de la tasa e indirectamente del riesgo sea cero, pues no plantean ninguna restricción a las defunciones observadas un año y para una edad y sexo determinado y éstas pueden ser cero como de hecho ocurre en muchos casos cuando la población es de pequeño tamaño.

Que el riesgo de muerte sea cero choca conceptualmente desde un punto de vista demográfico. Siempre existe un riesgo de morir. Otra cosa es que desde el punto de vista de la estadística descriptiva se observe que, en un determinado año, no fallecieron individuos de una determinada edad y el riesgo se compute como cero. Esto no significa que el riesgo de morir sea cero sino que el observado ese determinado año es cero. Naturalmente cualquier proyección tanto a tiempo real como a corto o largo plazo, no puede asumir en ningún caso un riesgo de muerte cero. Lo que sí puede es proporcionar una estimación para el riesgo y un intervalo de confianza para dicha estimación con un determinado nivel de confianza y que este intervalo incluya el cero.

El hecho de que aparezcan riesgos iguales a cero no es deseable aunque sí posible e incluso probable con el paso del tiempo y la disminución de la incidencia de la mortalidad a ciertas edades. Sobre todo desde el momento en que la población objeto de estudio no tiene suficiente tamaño para que las estimaciones en particular y la estadística en general elaborada a partir de la población sea consistente. La Estadística es la ciencia de los grandes números. Por ejemplo, estimar la esperanza de vida en una población de tamaño 1000 es computacionalmente posible, pero entre poco y nada consistente. Cuando tratamos con poblaciones de pequeño tamaño, factores circunstanciales que afecten a un individuo o a dos pueden influir dramáticamente en la estimación e incluso invalidarla. Una línea de investigación abierta es establecer el tamaño suficiente de una población para el cual es aceptable

la estimación. Se trata de un problema que depende de la incidencia de la mortalidad, pues cuanto más baja más difícil es estimarla y al mismo tiempo de la estructura de la población. Si nos desentendemos de esta cuestión nos haremos trampas a nosotros mismos al estudiar los resultados. Es como estimar la probabilidad de cara de una moneda con tres lanzamientos. El resultado de la estimación será 0, $1/3$, $2/3$ ó 1, o sea, nos equivocamos siempre.

Una primera aproximación al problema del tamaño mínimo para estimar la esperanza de vida es pensar en el cálculo del riesgo para una edad. Los valores mínimos del riesgo de muerte en España se alcanzan entre los 4 y los 12 años, entre los que se mueve por debajo de la diez milésima. Esto significa que necesitamos al menos 10.000 efectivos para empezar a tener valores fiables. Por 100 edades por 2 sexos queda una población sobre los 2 millones de personas para poder calcular una tabla de mortalidad con garantías. Aún así, en poblaciones como Madrid, con más de 6 millones de habitantes, no es excepcional que se obtenga riesgo de muerte cero en alguna edad u otro tipo de inconsistencias. Incluso los 46 millones del total de la población de España pueden ser insuficientes a tenor de la evolución de la mortalidad donde alcanza sus mínimos en torno a los 10 años de edad.

Otro factor a tener en cuenta es que a otras edades con riesgo de muerte más alto el tamaño de la muestra necesario para estimar el riesgo puede ser inferior. Sin embargo, las fuentes de información actuales en España desagregar la población como máximo por provincia. Con una fuente de información de población que permitiera discriminar la muestra por individuo se podría calcular una estimación más representativa y al mismo tiempo más reducida dando una población total inferior.

De entre las conclusiones de este capítulo hay una que tiene especial relevancia en relación con la estimación de la incidencia de la mortalidad. Casi todas las funciones biométricas de la tabla de mortalidad proporcionan medidas de la incidencia de muerte equivalentes. A partir de una función podemos deducir las demás. Podemos elegir la que queramos indistintamente. La diferencia principal aparece en cuanto a la base sobre la que se construye cada función. Se distinguen dos grupos. Por una parte están las funciones que son independientes por edad como la

tasa, el riesgo o las defunciones. Estas funciones se caracterizan por estar más expuestas a perturbaciones¹³. Como consecuencia, las estimaciones de estas funciones se caracterizan por ser inestables y menos consistentes. Frente a estas funciones, las funciones que se construyen recursivamente como la de supervivientes, la población estacionaria, el tiempo que queda por vivir o la esperanza de vida son más estables. Sus estimaciones son más consistentes porque la muestra que se toma como base para su estimación es más amplia y se va incrementando según se suba o baje en edad, de tal manera que la esperanza de vida al nacimiento parte de la observación de la incidencia de la mortalidad en toda la población mientras que el riesgo de muerte a los cero años solo tiene en cuenta la población de cero años. Con esta carta de presentación, la elección de la función a la hora de estimar y proyectar la mortalidad, a priori se decanta hacia las funciones recursivas como la esperanza de vida, sobre todo en poblaciones de pequeño tamaño.

Tradicionalmente se acepta sin demasiada discusión la población objeto de estudio pues está muy condicionado a las fuentes de información disponibles. Generalmente, la discriminación es exclusivamente política sin hacer consideraciones socioeconómicas, biológicas o de otro tipo. Por ejemplo, desde el momento en que se calcula un parámetro como la esperanza de vida al nacimiento de un varón para España se obtiene la estimación de la esperanza de vida para un individuo representativo de la población objeto de estudio. Se está considerando, por tanto, que todos los individuos de dicha población están sujetos a condiciones de exposición al riesgo de mortalidad comunes para todos los individuos de la población.

Ahora bien, la esperanza de vida de un individuo de una provincia y por tanto de España será la estimación obtenida para España. Aunque se pueda proporcionar una esperanza de vida observando sólo los individuos de una provincia, esta estimación será una estimación menos fiable desde el momento en que la muestra es más pequeña.

¹³ El lector interesado puede consultar Wrycza y Baudisch (2012) para una presentación teórica más exhaustiva de la relación entre el cambio en la intensidad de mortalidad y el cambio resultante en la esperanza de vida.

Aunque hay bibliografía que recoge la singularidad de la estimación en poblaciones de pequeño tamaño¹⁴, hay escaso consenso en el tamaño mínimo de la población que merezca consideración como entidad y sea digno de estudio. De hecho, el cálculo está condicionado a razones políticas y administrativas antes que a criterios estadísticos. El INE calcula la esperanza de vida por provincia, incluyendo a las dos Ciudades Autónomas Ceuta y Melilla. Independientemente del sentido que tenga la medición de cualquier magnitud social o de contabilidad nacional y regional, estas dos entidades administrativas oficialmente tienen su dato.

Es más, cuando estudiamos la esperanza de vida en una provincia al considerarla como un ente homogéneo, se desprecian las diferentes condiciones respecto a la exposición al riesgo que pueden existir dentro de esa población, por ejemplo, si se distingue entre población urbana y población rural. Las condiciones de la población urbana de Madrid pueden identificarse más con las de la población urbana de Barcelona que con las de la población rural de la misma provincia de Madrid.

El dato provincial está encubriendo el objetivo final que es, en realidad, conocer la esperanza de vida de un individuo representativo de unas condiciones socio-económicas particulares. Este dato teórico resulta imposible de conocer. Así que observamos una cierta población, como muestra, a partir de la cual estimarla. Un conjunto de individuos lo más representativo posible del individuo teórico objeto de interés. Una muestra en el entorno del individuo cuya situación socio-económica y biológica sean lo más parecidas a las suyas.

Una interesante e innovadora línea de investigación abierta, en la estimación de esperanza de vida de un individuo objeto de estudio, es la selección de una muestra más representativa eligiendo los elementos de esta muestra en función de distintas variables como la distancia al individuo teórico objeto de interés, su estatus económico, el tipo de municipio, rural o urbano, y ponderando según convenga. Con esta opción, cada punto tiene un valor de esperanza de vida y el agregado da una

¹⁴ El lector interesado puede acudir al artículo de López, Gutiérrez de Mesa y Rivera (2012) o también a Kanaroglou, Maoh, Newbold, Scott y Páez (2009), Gabriel (1992) o Cohen (1977), por ejemplo.

superficie continua de esperanza de vida a semejanza de un mapa de relieve con curvas de nivel.

Esta alternativa requiere una discriminación de la muestra de la población objeto de estudio que actualmente solo se puede hacer con estudios demográficos longitudinales donde se dispone de registros de cada individuo con la exposición al riesgo y además las variables a partir de las cuales se va a discriminar la muestra. El padrón municipal es un posible punto de partida para un estudio que requiere un tratamiento particular de las fuentes de información. La Base de Datos Longitudinal de Población de Andalucía es un importante avance en este sentido (véase por ejemplo Viciano (2010)).

Discriminar la población objeto de estudio por razones políticas es una opción desde el momento en que constitucional y legalmente se considera a todos los individuos de una Nación con los mismos derechos y deberes económicos, sociales, educativos, etc. Sobre todo es la única opción en España porque sólo se dispone de información desagregada por provincias.

Al margen de las anteriores consideraciones, a continuación, teniendo en cuenta todas las metodologías revisadas, se presenta una alternativa metodológica que es en esencia una extensión de la metodología del INE (2013). Se abre por tanto una discusión sobre alternativas metodológicas en las tablas de mortalidad de España con el objetivo de apurar todas las posibilidades que ofrecen las fuentes de información disponibles.

En primer lugar se prescinde de la fórmula 2.42 y se extiende la tabla a todas las edades posibles mientras los datos sean consistentes. Esta opción se justifica por el siguiente razonamiento empírico.

En la tabla 2.2 figuran los datos del Movimiento Natural de Población de 2012, INE (2012a). De acuerdo con estos datos¹⁵ fallecieron 10 niñas de 10 años y 3726 mujeres de 100 o más años siendo las mujeres fallecidas con 99 un total de 1758. De

¹⁵ A pesar de las dudas que generan los registros a edades tan avanzadas, se trata de datos que vienen avalados por el registro civil. Además a partir de los 107 años se pone en marcha un proceso de validación (Jeune y Vaupel, 1999).

acuerdo con la metodología del INE (2013) se calcula la tabla de mortalidad para el total nacional por edades simples que incluye la edad 10. La frecuencia absoluta de la clasificación de 100 y más años multiplica en casi 400 veces la frecuencia absoluta para la edad 10. En cambio la tabla de mortalidad se calcula para 10 años despreciando una rica información en edades superiores a 100 para estudios de longevidad.

Tabla 2.2 Fallecidos de 2012 clasificados por sexo

| Edad | Varones | Mujeres | Total |
|-----------------------|---------|---------|-------|
| 100 años | 375 | 1.266 | 1.641 |
| 101 años | 208 | 829 | 1.037 |
| 102 años | 127 | 503 | 630 |
| 103 años | 70 | 335 | 405 |
| 104 años | 34 | 186 | 220 |
| 105 años | 16 | 114 | 130 |
| 106 años | 18 | 57 | 75 |
| 107 años | 6 | 37 | 43 |
| 108 años | 2 | 11 | 13 |
| 109 años | 2 | 10 | 12 |
| 110 y más años | 1 | 3 | 4 |

Fuente: Elaboración propia

Si calculamos el riesgo de muerte de las mujeres a la edad de 10 años con un total de defunciones de 10, esta cantidad se alcanza al menos hasta 105 años. A partir de esta edad los datos fluctúan mucho y pierden consistencia.

Por tanto, se propone prolongar la tabla de mortalidad para el total nacional hasta los 105 años.

Por otra parte, decir que el riesgo de muerte para el intervalo abierto es uno no proporciona información relevante más allá de garantizar que la generación ficticia se extingue, como por otra parte ya se sabía. Se propone estimar la esperanza de vida a los w años por la media del tiempo vivido por los individuos que fallecen en el intervalo de w y más años por la expresión:

$$a(t, w, s) = \frac{\sum_{i=1}^{D(t,w,s)} a(t, w, s, i)}{D(t, w, s)}, \quad (2.45)$$

donde $a(t, w, s, i)$ se calcula de acuerdo con las metodologías del INE (2013) y INE (2009) teniendo en cuenta que cuando la edad sea superior a w años se suma los años transcurridos desde que se cumplieron los w años.

Esta estimación proporciona una información más acertada sobre el tiempo que se puede esperar que viva una persona en el último intervalo de edad simple consistentemente calculado.

Este valor ha de observarse con precaución porque mezcla todas edades sin tener en cuenta la diferente distribución de las defunciones para cada edad simple ya que frente a la filosofía de una tabla de mortalidad que mide la incidencia de la mortalidad con independencia del sexo y la edad, con la fórmula 2.45 hay dependencia en la edad.

Por otra parte, se presenta a continuación un resultado observado para la metodología del INE. Se trata de una discrepancia entre metodologías aunque con escasa influencia en los resultados. Donde mejor se aprecia es el año 2002 cuando la diferencia entre las poblaciones ePOBa e intercensales es más pequeña. En particular en la tasa de mortalidad¹⁶ de las edades avanzadas donde el valor para la metodología INE(2013) es inferior al valor para la metodología INE(2009). La última publicación que reconstruye la serie de tablas de mortalidad intercensales desde 2002 a 2011 se calcula con la metodología INE (2013), aunque el resto de años mantiene la metodología INE (2009).

Así pues, de acuerdo con la metodología INE (2009) la estimación de la tasa en el año t a la edad x presenta un sesgo. El valor insesgado de la exposición al riesgo de la población se calcula como suma de la exposición para la generación g y la generación $g-1$:

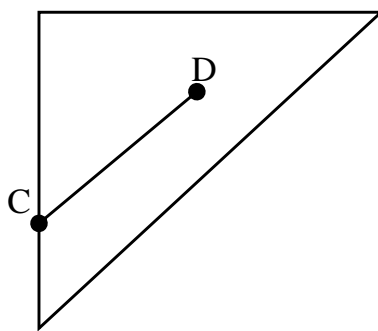
Para la generación g , descrita en el diagrama de Lexis por el triangulo inferior correspondiente en la Figura 2.18, la exposición al riesgo se calcula de forma análoga a la metodología por el INE (2009):

¹⁶ Tablas de mortalidad del INE. www.ine.es

$$1/2 \cdot P(t, x, g, s) + \sum_i b(t, x, g, s, i) \quad (2.46)$$

Para el caso de la generación $g-1$, representada en la Figura 2.18 por el triángulo superior, la exposición al riesgo de los individuos de la generación $g-1$ que llegan al 1 de enero del año t y sobreviven hasta cumplir $x+1$ años es $1/2 \cdot (P(t, x, g-1, s) - D(t, x, g-1, s))$. A esta exposición hay que añadirle el tiempo vivido por los fallecidos $D(t, x, g-1, s)$.

Figura 2.18. El valor del segmento b' se mide como D (fecha de defunción) menos C (fecha de 1 de enero del año t).



Fuente: Elaboración propia

Es decir la suma de los segmentos CD que comienzan en el punto C a 1 de enero del año t y terminan en el punto D , fecha de fallecimiento. Sea $b'(t, x, g-1, s, i)$ el tiempo en años para el individuo i del sexo s , generación $g-1$ que fallecen con edad x el año t descrito en la Figura 9 por el segmento CD . La exposición al riesgo de la población el año t con edad x de la generación $g-1$, descrito en el diagrama de Lexis por un triángulo superior como en la Figura 2.17, es:

$$1/2 \cdot (P(t, x, g-1, s) - D(t, x, g-1, s)) + \sum_i b'(t, x, g-1, s, i) \quad (2.47)$$

Con lo que la exposición al riesgo insesgada, denominador de la tasa de mortalidad en el año t con edad x bajo las hipótesis asumidas por la metodología del INE (2013), es:

$$1/2 \cdot (P(t, x, g - 1, s) - D(t, x, g - 1, s)) + \sum_i b'(t, x, g - 1, s, i) + 1/2 \cdot P(t, x, g, s) + \sum_i b(t, x, g, s, i)$$

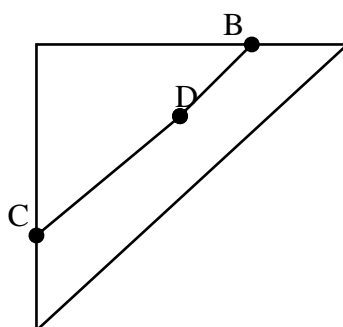
Se tiene que la diferencia entre la exposición al riesgo de la metodología INE (2009) $E_o(t, x)$ y la exposición al riesgo con la metodología INE (2013) $E_c(t, x)$ es:

$$\begin{aligned} E_o(t, x) - E_c(t, x) &= 1/2 \cdot P(t, x, g - 1, s) + \sum_i b(t, x, g - 1, s, i) + 1/2 \cdot P(t, x, g, s) + \sum_i b(t, x, g, s, i) \\ &- [1/2 \cdot (P(t, x, g - 1, s) - D(t, x, g - 1, s)) + \sum_i b'(t, x, g - 1, s, i) + 1/2 \cdot P(t, x, g, s) + \sum_i b(t, x, g, s, i)] = \\ &= 1/2 \cdot D(t, x, g - 1, s) - \sum_i b'(t, x, g - 1, s, i) + \sum_i b(t, x, g - 1, s, i) = \\ &= 1/2 \cdot D(t, x, g - 1, s) + \sum_i (-b'(t, x, g - 1, s, i) + b(t, x, g - 1, s, i)) \end{aligned} \tag{2.48}$$

Como se puede observar en la Figura 2.19, $-b'(t, x, g - 1, s, i) + b(t, x, g - 1, s, i)$ es la suma del segmento BD más el segmento DC, i.e., el segmento BC, el cual si el corte de las líneas de vida con el 1 de enero del año t es uniforme en media vale $1/2$. En la práctica no difiere significativamente de $1/2$ que es la hipótesis de la metodología del INE (2009) y la diferencia $E_o(t, x) - E_c(t, x)$ es cero.

El segmento CD de la Figura 2.19 representa la línea de vida de un individuo i donde C es el punto en que la línea de vida pasa por el 1 de enero del año t y D es el fallecimiento y mide $b'(t, x, g - 1, s, i)$. El segmento BD representa la línea de vida del individuo i, donde B es la fecha del cumpleaños que no llega a celebrar y mide $b(t, x, g - 1, s, i)$.

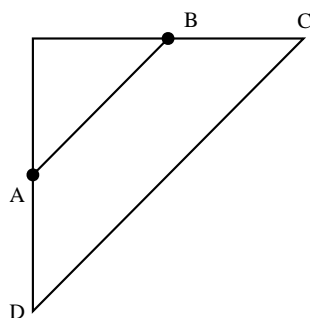
Figura 2.19. Línea de vida del individuo i de la generación g-1 el año t con edad x.



Fuente: Elaboración propia

La diferencia aumenta con la edad, porque se concentran más defunciones, y por tanto más segmentos $-b'(t, x, g - l, s, i) + b(t, x, g - l, s, i)$, en la región ABCD de la Figura 2.20.

Figura 2.20. Triángulo superior en el diagrama de Lexis.



Fuente: Elaboración propia

Por tanto se tiene como resultado que para las edades altas la metodología INE (2009) subestima la exposición al riesgo, de lo que se deduce que sobreestima la tasa de mortalidad y el riesgo y subestima la esperanza de vida.

Este resultado no se refleja en los datos de la serie revisada publicados por el INE (2013) que reflejan como se ve en la tabla 2.3 una aparente subestimación de la esperanza de vida con la metodología INE 2009 frente a la metodología 2013 en los primeros años que se reduce según pasa el tiempo para pasar a una subestimación en los últimos años. Esta diferencia en realidad no puede achacarse a la metodología pues el factor dominante es el cambio en los stock de población intercensal. El censo de 2011 da una población de 46.818.216 residentes a 1 de enero de 2012¹⁷, significativamente superior a 46.196.278 residentes, cifra estimada por ePOBa¹⁸ para esa fecha. La revisión incrementa significativamente la población, 621.938 residentes, un 13,46% más. La diferencia se reduce progresivamente según se retrocede en el tiempo y nos acercamos al censo de 2001 en que coinciden los stock a 1 de enero. La

¹⁷ A partir de los datos del censo a 1 de noviembre de 2011: 46.815.916 residentes. [www.ine.es]

¹⁸ Fuente: Datos de la estimación de población actual www.ine.es.

consecuencia directa es un aumento de la esperanza de vida al repartirse el mismo número de defunciones entre más población.

Tabla 2.3. Esperanza de vida al nacimiento con la metodología 2013 y la metodología 2009.

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|--------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Metodología 2009 Mujeres | 83,144 | 82,992 | 83,596 | 83,548 | 84,156 | 84,106 | 84,271 | 84,564 | 84,910 | 84,973 |
| Metodología 2013 Mujeres | 83,141 | 82,992 | 83,581 | 83,539 | 84,157 | 84,142 | 84,338 | 84,664 | 85,050 | 85,161 |
| Diferencia Mujeres M2013-M2009 | -0,003 | 0,000 | -0,015 | -0,009 | 0,001 | 0,036 | 0,068 | 0,100 | 0,140 | 0,188 |
| Metodología 2009 Varones | 76,402 | 76,424 | 76,996 | 77,028 | 77,716 | 77,766 | 78,174 | 78,545 | 78,939 | 79,157 |
| Metodología 2013 Varones | 76,383 | 76,417 | 76,983 | 77,020 | 77,709 | 77,787 | 78,228 | 78,631 | 79,058 | 79,317 |
| Diferencia Varones M2013-M2009 | -0,019 | -0,008 | -0,013 | -0,008 | -0,007 | 0,022 | 0,053 | 0,085 | 0,119 | 0,160 |

Fuente: Elaboración propia a partir de datos del INE

La alternativa metodológica propuesta logra una mejor aproximación al estudio de los superlongevos, población de 100 y más años, etc.; objeto de especial interés en la actualidad por el aumento de la esperanza de vida. Además proporciona una estimación de la esperanza de vida a la edad de cierre de la tabla de mortalidad de la que carece la metodología (2013).

La alternativa a la metodología INE (2013) explota hasta sus últimas posibilidades las actuales fuentes de información disponibles. Son las propias fuentes de información las que limitan las posibilidades. El ideal es un registro a modo de censo continuo y es esta línea de trabajo la que está explorando el INE desde 1999. En un primer proyecto denominado Estudio Demográfico Longitudinal¹⁹, en adelante EDL, que dirigió Eduardo Teijeiro hasta 2009 y que continuó Carmen Teijeiro hasta 2012, fecha en que el INE decidió abandonar el proyecto. Este proyecto pretendía reconstruir las líneas de vida de las personas partiendo del censo de 2001 aprovechando el padrón municipal. A pesar del trabajo realizado no se pudo validar porque el censo de 2011 no fue exhaustivo. Por otra parte el EDL, aunque supone una importante evolución, recoge la población en puntos discretos cada mes sin llegar al continuo temporal.

¹⁹ El lector interesado en el proyecto puede consultar la ficha histórica que figura en el Inventario de Operaciones Estadísticas del INE. http://www.ine.es/ss/Satellite?L=es_ES&c=TFichaHistorica_C&cid=1259930831802&p=1254735089786&pagename=IOEhist%2FIOEhistLayout¶m1=normal

El INE abandonó esta línea de trabajo y apostó por un ambicioso proyecto denominado Sistema de Información Geográfica de Población y Territorio, en adelante SIG_PT, que es el que actualmente está en marcha. Un proyecto que comenzó llevándose desde varias administraciones y que finalmente se integró en el INE. Se trata de un sistema en línea orientado a servicios para la gestión integrada de las variaciones de población y de territorio de los ciudadanos residentes en España y de los españoles residentes en el extranjero, de forma que estén permanentemente actualizados.

El objetivo de este proyecto es ligar cada individuo a una posición, es decir, georreferenciar a los residentes y completar la información con los datos del padrón, administrados municipalmente y con una gestión centralizada desde el INE como ahora pero con el plus de ser en tiempo real. Esta opción requiere la implementación de una nueva aplicación actualmente en construcción y que se espera dé sus primeros frutos en los próximos dos años. Lo deseable sería que el SIG_PT sea una realidad al menos para el 2021, año de censo, y que por tanto evite una parte importante de la operación estadística más costosa que realiza el INE.

Indirectamente cuando el SIG_PT funcione la elaboración de las tablas de mortalidad pasará de ser una estimación a un cálculo contable. Mientras tanto, el INE dispone de un buen instrumento ya consolidado para la medición de la mortalidad y que se ha descrito en el apartado 2.6. El trabajo en el siguiente capítulo se centra en aplicar la herramienta al caso de España, tanto para el total nacional como a nivel territorial, a lo largo de dos décadas.

El resultado es un importante volumen de datos. Una rica base de datos que proporciona interesante información. La observación de esta base de datos abre nuevas líneas de investigación. En particular una de las líneas abiertas más prometedoras, y sobre la que el trabajo futuro se orienta, es analizar la particular evolución de la nube de series de los riesgos de mortalidad de las provincias en torno al riesgo para el total nacional.

Capítulo 3

El análisis de las tablas de mortalidad de España y sus provincias en el periodo 1981-2008

- 3.1. Introducción
- 3.2. Tablas de mortalidad de España
- 3.3. Tablas de mortalidad de las provincias españolas
 - 3.3.1. El riesgo de muerte
 - 3.3.2. La esperanza de vida
- 3.4. Clasificación de las provincias por esperanza de vida
 - 3.4.1. Análisis por año
 - 3.4.2. Análisis combinado
- 3.5. Conclusiones

3.1. Introducción

Este capítulo se ocupa de un análisis descriptivo de las funciones de la tabla de mortalidad de España y sus provincias a lo largo de los 18 años observados que comprenden el periodo desde 1991 hasta 2008. Los resultados presentados se han elaborado a partir de los datos observados en España y sus provincias de las tablas de mortalidad tanto para varones como para mujeres de acuerdo con la metodología oficial del INE (2009). Para España se calcula por edades simples hasta 100 años y más y por provincias por edades simples hasta 95 y más excepto para Ceuta y Melilla, cuyas tablas se cierran en 90 años y más.

Los datos se han procesado utilizando un programa de SAS, en las siguientes etapas:

1. Se toma como datos de entrada las poblaciones a 1 de enero del año t y el año $t+1$ y las defunciones del año t . Las poblaciones se importan desde una hoja de cálculo de Excel a partir de los datos de la web del INE. Para el cálculo de las tablas de mortalidad se necesitan las poblaciones a 1 de enero de cada año. La población se obtiene de acuerdo con la metodología de las *estimaciones de población actual* del INE (2008) que se calculan recursivamente cada año a partir del anterior por el método de las componentes. La última actualización de la metodología INE (2012b) que se aplica desde octubre de 2012 incorpora información mensualmente. Tanto a nivel nacional como provincial están desagregadas por sexo y edad simple hasta 100 y más años. El nivel provincial es el máximo nivel de desagregación territorial y cierra a 100 años, a diferencia de las tablas de mortalidad que cierran en las provincias a menor edad para asegurar la consistencia de los datos.

Figura 3.1. Diseño de registro del fichero con los datos de defunciones del MNP.

| I.N.E. | | DESCRIPCION DE REGISTRO | | | | PAGINA 1/1 | |
|------------------|--------------|--|----------|----------|--------------|------------|--|
| ESTADISTICA : | CODIGO = 321 | NOMBRE = REGISTRO FINAL DE DEFUNCIONES DESDE 1999 EN ADELANTE | | | | | |
| FICHERO LOGICO : | NOMBRE = | ESTRUC.= UNICO FIJO | | | | | |
| REGISTRO : | TIPO = | LONGITUD = 85 | | | | | |
| | | CONTENIDO = MOVIMIENTO NATURAL DE LA POBLACION: DEFUNCIONES (A PARTIR DEL AÑO 1.999) | | | | | |
| NIVEL | CAMPO | REPET | COMIENZO | LONGITUD | TIPO DE DATO | CLAVE | DESCRIPCION DEL CAMPO |
| 1 | NUMREC | 1 | 1 | 2 | CHAR(2) | | NUMERO REGISTRO CIVIL |
| 1 | MUNDEF | 1 | 3 | 3 | PIC'999' | | MUNICIPIO DONDE SE INSCRIBIO LA DEFUNCION |
| 1 | PROVDEF | 1 | 6 | 2 | PIC'99' | | PROVINCIA DONDE SE INSCRIBIO LA DEFUNCION |
| 1 | TOMO | 1 | 8 | 4 | CHAR(4) | | TOMO |
| 1 | LIBRO | 1 | 12 | 3 | CHAR(3) | | LIBRO |
| 1 | PGNA | 1 | 15 | 3 | CHAR(3) | | PAGINA |
| 1 | VUELTA | 1 | 18 | 1 | CHAR(1) | | VUELTA |
| 1 | FECHANA | 1 | 19 | 0 | | | FECHA DE NACIMIENTO |
| 2 | DIAN | 1 | 19 | 2 | PIC'99' | | DIA DEL NACIMIENTO |
| 2 | MESN | 1 | 21 | 2 | PIC'99' | | MES DEL NACIMIENTO |
| 2 | ANON | 1 | 23 | 4 | PIC'9999' | | AÑO DEL NACIMIENTO |
| 1 | PRONN | 1 | 27 | 2 | CHAR(2) | | PROVINCIA DE NACIMIENTO |
| 1 | MUNN | 1 | 29 | 3 | CHAR(3) | | MUNICIPIO DE NACIMIENTO |
| 1 | SEXO | 1 | 32 | 1 | PIC'9' | | SEXO |
| 1 | ECIV | 1 | 33 | 1 | PIC'9' | | ESTADO CIVIL |
| 1 | NACIO | 1 | 34 | 3 | CHAR(3) | | NACIONALIDAD |
| 1 | PROF | 1 | 37 | 2 | PIC'99' | | PROFESION |
| 1 | MUNIRES | 1 | 39 | 3 | CHAR(3) | | MUNICIPIO DE RESIDENCIA |
| 1 | PROVRES | 1 | 42 | 2 | PIC'99' | | PROVINCIA DE RESIDENCIA |
| 1 | FECHADEF | 1 | 44 | 0 | | | FECHA DE LA DEFUNCION |
| 2 | DIADDEF | 1 | 44 | 2 | PIC'99' | | DIA DE LA DEFUNCION |
| 2 | MESEDEF | 1 | 46 | 2 | PIC'99' | | MES DE LA DEFUNCION |
| 2 | AÑODEF | 1 | 48 | 4 | PIC'9999' | | AÑO DE LA DEFUNCION |
| 1 | CAUSA | 1 | 52 | 0 | | | CAUSA DE LA DEFUNCION |
| 2 | CAUSAL | 1 | 52 | 1 | CHAR(1) | | CARACTER |
| 2 | CAUSA23 | 1 | 53 | 2 | PIC'99' | | NUMERICO |
| 2 | CAUSA4 | 1 | 55 | 1 | PIC'9' | | NUMERICO O BLANCO |
| 1 | AÑOSC | 1 | 56 | 3 | PIC'999' | | AÑOS CUMPLIDOS |
| 1 | MESC | 1 | 59 | 2 | PIC'99' | | MESES CUMPLIDOS |
| 1 | DIAC | 1 | 61 | 2 | PIC'99' | | DIAS CUMPLIDOS |
| 1 | TAMAMUNI | 1 | 63 | 1 | PIC'9' | | TAMAÑO MUNICIPIO INSCRIPCION |
| 1 | TAMAMUNR | 1 | 64 | 1 | PIC'9' | | TAMAÑO MUNICIPIO RESIDENCIA |
| 1 | CAUSAR | 1 | 65 | 3 | PIC'999' | | CAUSA DE MORTALIDAD REDUCIDA (102 GRUPOS) |
| 1 | CAUSAP | 1 | 68 | 2 | PIC'99' | | CAUSA DE MORTALIDAD PERINATAL (93 GRUPOS) |
| 1 | CAUSAINF | 1 | 70 | 2 | PIC'99' | | CAUSA DE MORTALIDAD INFANTIL (54 grupos) PARA MENORES DE 1 AÑO |
| 1 | NBOL | 1 | 72 | 8 | PIC'9(8)' | | NUMERO DE BOLETIN |
| 1 | ANOR | 1 | 80 | 4 | PIC'9999' | | AÑO RECEPCION |
| 1 | MESR | 1 | 84 | 2 | PIC'99' | | MES RECEPCION |

Fuente: INE

Las defunciones se toman como microdatos desde un fichero de texto plano con el diseño de registro que se puede observar en la Figura 3.1.

2. Se calculan las exposiciones al riesgo de los fallecidos dependiendo de la generación a la que pertenezcan y el tiempo vivido por el fallecido. Cuando la defunción ocurre en el triángulo inferior del diagrama de Lexis se calcula el dato directamente. Cuando la defunción pertenece al triángulo superior se toma como tiempo vivido el último año de vida incluyendo el trozo que transcurre en el año anterior.
3. Se suman las defunciones, las exposiciones al riesgo y se calcula la media del tiempo vivido el último año de vida con el procedimiento *summary* de SAS.
4. Se unen las poblaciones por una parte y las defunciones, exposición al riesgo y tiempo vivido el último año de vida en un conjunto data.
5. Se calcula la tasa de mortalidad y con el tiempo vivido el último año de vida, el riesgo de muerte para cada edad a partir de los parámetros del conjunto data.
6. Se toman la tasa, los años vividos el último año de vida y el riesgo y se construye una matriz con el procedimiento *ilm* con las dimensiones de la tabla de mortalidad.

7. Se calculan las funciones biométricas de la tabla de mortalidad. Primero la función de supervivientes junto con la de defunciones, de forma recursivamente. Luego la población estacionaria, el tiempo que queda por vivir y la esperanza de vida.

Este proceso se repite para los 18 años, de 1991 a 2008, las 52 provincias y los 2 sexos dando lugar un importante volumen de datos, que llega hasta los 1.434.816¹ a los que hay que añadir los 45.248² datos de 28 años para el total nacional que hacen un total de 1.480.064 datos como resultado de procesar más de 9 millones de registros de fallecidos.

Con la perspectiva de mejorar el conocimiento de la evolución de las funciones biométricas se presenta un innovador análisis descriptivo utilizando gráficos de las superficies de mortalidad que permiten observar la evolución en el tiempo, por edad y por generación para las tablas de mortalidad de España.

Se trata de gráficos a semejanza de los mapas de relieve, en los que las curvas de nivel separan los colores que delimitan regiones, y donde la función representada toma un valor del rango indicado en la leyenda. Estos gráficos están generados con lenguaje R a partir de código facilitado por Carlo di Camarda.

Por otra parte, y para conocer más en profundidad el comportamiento de la intensidad de la mortalidad provincial, se calculan las tablas de mortalidad para cada provincia por edades simples. Una pequeña parte, aunque representativa, de los resultados, se expone en mapas ilustrando la dispersión por provincias. Estos gráficos están generados con el *Enterprise Manager* de SAS. Se concluye con una discusión sobre los resultados obtenidos.

Este capítulo consta de cinco apartados:

- En el segundo apartado se presentan un conjunto de gráficos con las superficies de mortalidad para cada una de las funciones biométricas de España junto con índices de interés demográfico que se derivan directamente de las funciones de la tabla de mortalidad.

¹ 2 sexos x 8 funciones x18 años (50 provincias x 96 edades + 2 ciudades autónomas x 91 edades)

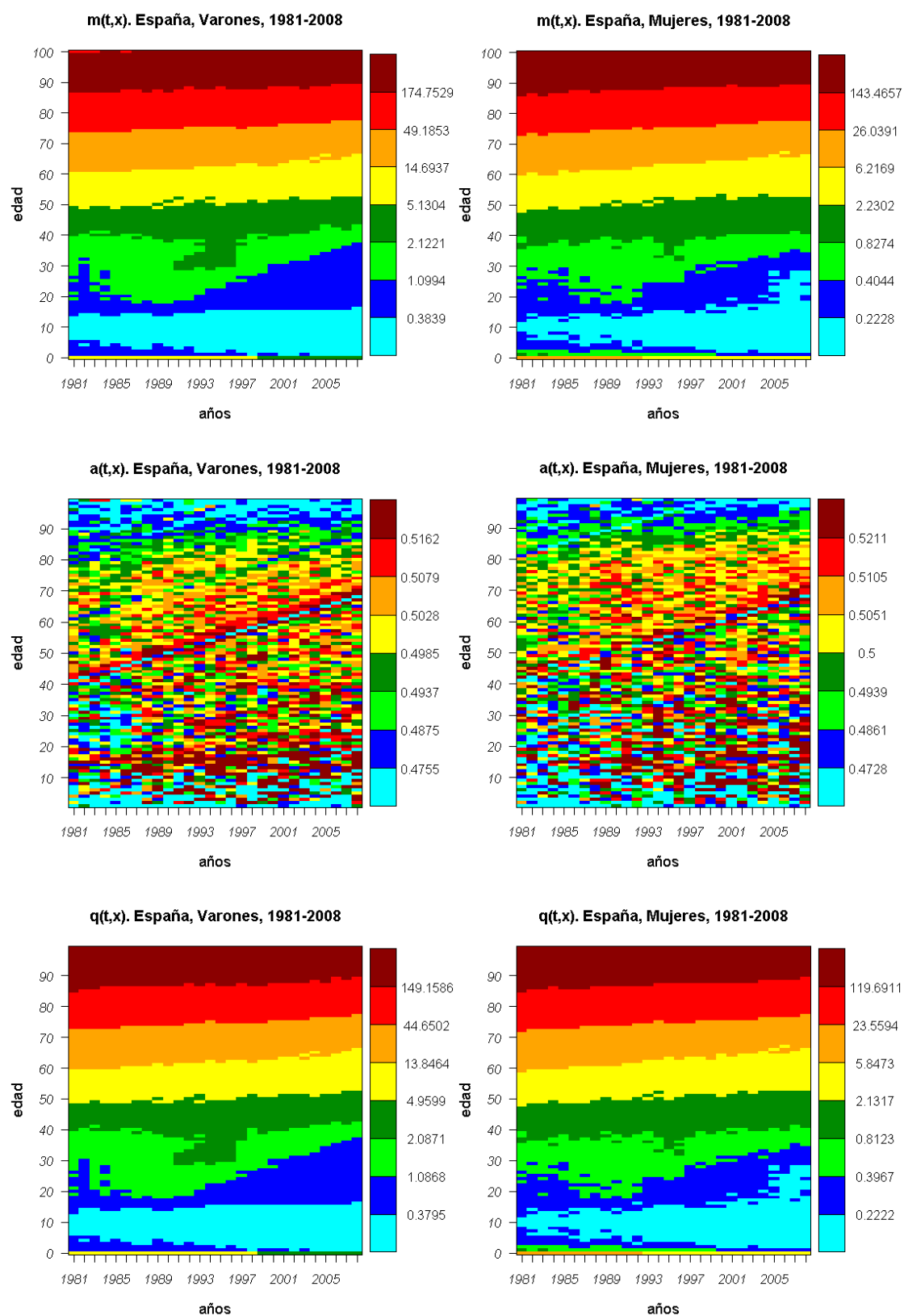
² 2 sexos x 8 funciones x28 años (Total Nacional x 101 edades)

- Seguidamente, en el tercer apartado, se analiza la evolución de la mortalidad en las provincias con gráficas y mapas de coropletas en dos subapartados, uno para riesgo de muerte y otro para la esperanza de vida.
- A continuación, en el cuarto apartado se clasifican las provincias por su mortalidad, utilizando como indicador la esperanza de vida a las distintas edades.
- Finalmente, el último apartado recoge las conclusiones del capítulo, incluyendo una discusión sobre la naturaleza del comportamiento de las variables riesgo y tasa de las provincias en particular respecto del total nacional. Incluye un subapartado para una clasificación por año y otro subapartado para una clasificación combinada a lo largo del tiempo.

3.2. Tablas de mortalidad de España

En las gráficas de la Figura 3.2 y siguientes se representan las funciones biométricas de la tabla de mortalidad observadas desde 1981 hasta 2008. Los años posteriores observados, de 2009 a 2011, se utilizarán para verificar la bondad de posibles ajustes. Todas las funciones biométricas se han calculado para el rango de edades de 0 a 100 años y más. Para cada cuadrícula del plano cartesiano, con el tiempo t en años en el eje de abscisas y la edad x en años en el eje de ordenadas, las funciones biométricas toman un valor que se representa por el color correspondiente al rango indicado en la leyenda. El tiempo vivido $a(t, x)$ no se representa para la edad 100 porque no es comparable con el resto de valores para otras edades y su inclusión distorsiona la percepción del gráfico. El riesgo $q(t, x)$ tampoco incluye los valores para la edad 100 porque ya se sabe su valor es 1 y no aporta información relevante. La tasa y el riesgo se expresan en milésimas y el tiempo vivido en años. Los supervivientes y las defunciones en individuos. La población estacionaria y el total de años vividos en años-persona y la esperanza de vida en años.

Figura 3.2. Superficies de mortalidad para las funciones $m(t, x)$, $a(t, x)$ y $q(t, x)$ de 1981 a 2008. Hasta 100 años de edad. Izquierda varones, derecha mujeres.



Fuente: Elaboración propia a partir de datos del INE

El estudio de los resultados permite observar distintas características sobre la evolución de las distintas funciones al combinar los parámetros de tiempo t y edad x para dar lugar a una superficie. En una primera aproximación se aprecia un alto grado de contigüidad en las superficies tanto respecto al tiempo como a la edad en todas las funciones biométricas.

En un estudio más detallado de la Figura 3.2, en particular de la gráfica de los años vividos el último de vida, llaman la atención unas líneas transversales sobre la superficie. Estas líneas siguen las generaciones terminadas en cero a partir de 1940 y anteriores tanto para varones como para mujeres.

La función $a(t, x)$ de la Figura 3.2 no incluye la edad 0, con valores significativamente inferiores, porque distorsiona la representación de la superficie. Se distinguen distintas regiones en relación con la edad. En las primeras edades se observan valores inferiores al medio año. Pasan a ser superiores al medio año en la adolescencia y juventud y se estabiliza en torno al medio año al aumentar la edad. A partir de los 90 años comienza a decrecer y cada vez son menores según aumenta la edad.

Para las generaciones que terminan en cero se observan valores menores al resto de generaciones adyacentes. En particular para la generación de 1940, 1920 y 1900 tanto para varones como para mujeres. La explicación de este fenómeno es una línea de investigación abierta. Se descartan influencias de error de redondeo del censo descrito en Goerlich (2008) porque la fuente de información del parámetro es el MNP. La causa tiene que venir del protocolo de registro en el MNP. Se plantea como hipótesis que cada año se lleve a cabo un proceso de depuración sistemática en el que los fallecidos para los que se desconozca la fecha de nacimiento y la generación se les impute un fallecimiento próximo al comienzo del año y se les asigne una generación terminada en cero por nuestra tendencia natural a redondear y sobre todo a las generaciones de 1900, 1920, 1940, 1960, etc. Estas defunciones sin clasificar serán las menos y aparecerán con más frecuencia cuando hay más fallecimientos, i.e., en edades avanzadas. Este criterio explicaría que sistemáticamente estas generaciones presenten

un $a(t,x)$ inferior a generaciones contiguas. El fenómeno se corrige para generaciones posteriores a 1970, en sintonía con la mejora y aumento de fiabilidad de los registros.

Apoya esta tesis que la tasa de mortalidad de estas generaciones terminadas en cero sea ligeramente superior que la tasa de las generaciones contiguas. Este hecho se puede apreciar en particular en la gráfica para $m(t,x)$ de los varones de la Figura 3.2 donde aparecen unas marcas de tendencia en la generación de nacidos en 1940, con valores mayores que sus generaciones contiguas. Esto indica que el número de defunciones es mayor. Causa cuya única explicación es achacable a un sesgo en el criterio que impute defunciones a esa generación terminada en cero que a otras.

En la Figura 3.2 se aprecia como la superficie del riesgo toma valores menores que la tasa como cabía esperar por definición. Por otra parte tanto para varones como para mujeres se observa que la forma de las superficies determinadas por la tasa y el riesgo son semejantes.

La evolución es regular y con contigüidad, sin cambios bruscos. Siempre con valores superiores en los varones. Se observa un repunte de la mortalidad en los varones de edades comprendidas entre los 30 y los 40 años a principios de los años noventa.

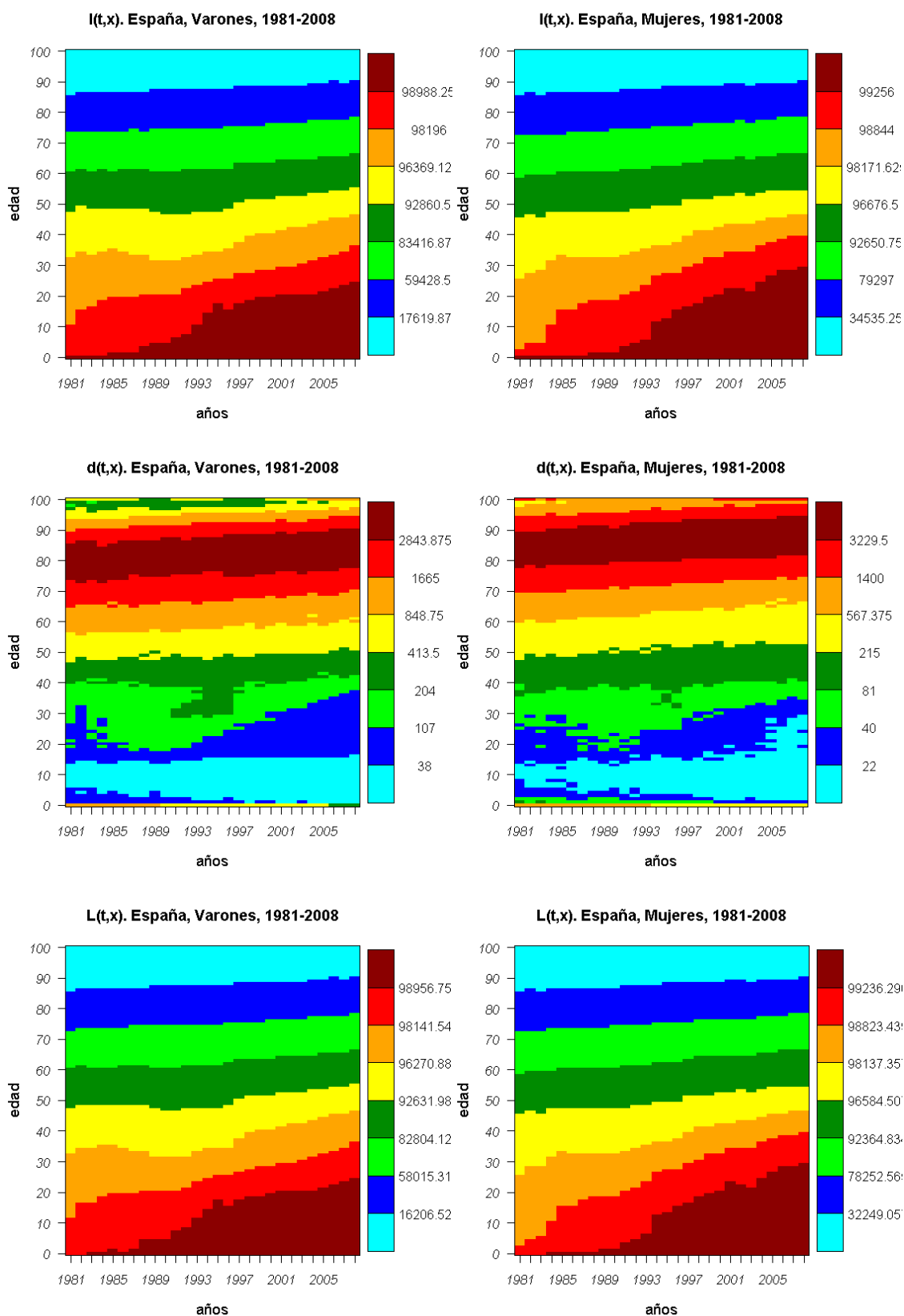
En la Figura 3.3 se observa que la evolución de los supervivientes y la población estacionaria son semejantes. Hay un mayor crecimiento de la superficie en torno al año 1990, que es más brusco para la población estable. Las dos superficies reflejan un crecimiento en el tiempo aunque con diferencias entre sexos. En torno a los 90 años todavía sobrevive un tercio de los efectivos iniciales mientras que para los varones solo sobrevive un sexto.

El comportamiento de estas funciones es más suave que el de las funciones tasa y riesgo, factor que favorece su estimación.

La moda de las defunciones teóricas se va desplazando con el tiempo a edades superiores. Tanto para varones como para mujeres crece unos dos meses por año desde los 80 años en los varones y es unos 5 años superior en las mujeres, diferencia que se conserva a lo largo del tiempo.

También las defunciones teóricas son una alternativa a considerar a la hora de modelizar. Esta función tiene la propiedad de que su agregación da siempre los 100.000 individuos de la cohorte ficticia que se exponen al fenómeno demográfico.

Figura 3.3. Superficies de mortalidad para las funciones $l(t,x)$, $d(t,x)$ y $L(t,x)$ de 1981 a 2008. Hasta 100 años de edad. Izquierda varones, derecha mujeres.

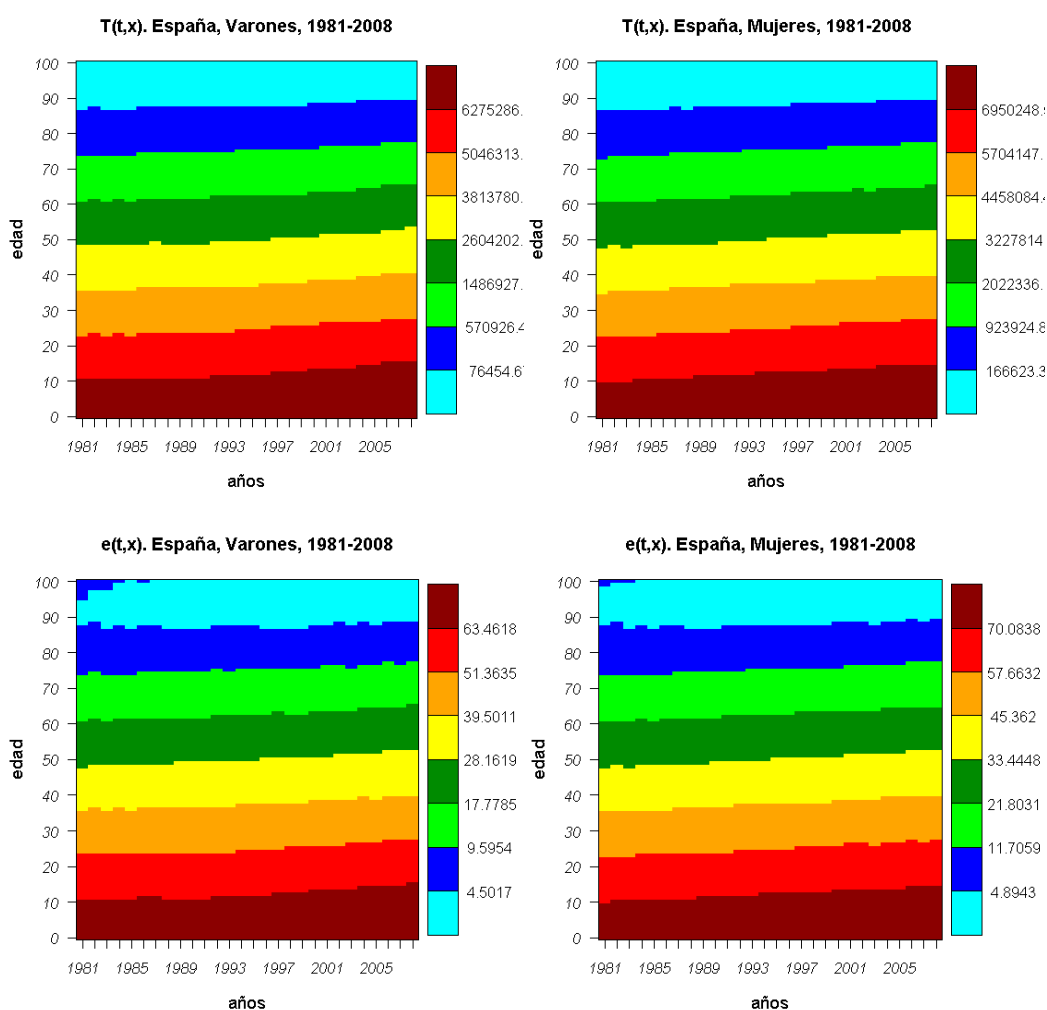


Fuente: Elaboración propia a partir de datos del INE

Uno de los principales motivos para el estudio de las tablas de mortalidad es en realidad la utilización de parámetros como el riesgo de muerte en la proyección de la población. Otra línea de investigación abierta es utilizar directamente los eventos demográficos, es decir, las defunciones reales y ajustarlas directamente sin procesos intermedios a través de las funciones biométricas.

Más suave todavía es el comportamiento de las superficies que representan los años vividos totales y la esperanza de vida. Este es un factor importante a tener en cuenta a la hora de estudiar el comportamiento de la mortalidad frente a otras variables, como la tasa y el riesgo, cuyas fluctuaciones distorsionan y dificultan la apreciación de la evolución, sobre todo en poblaciones de pequeño tamaño como veremos en el apartado 3.3.

Figura 3.4. Superficies de mortalidad para las funciones $T(t,x)$ y $e(t,x)$ de 1981 a 2008. Hasta 100 años de edad. Izquierda varones, derecha mujeres.

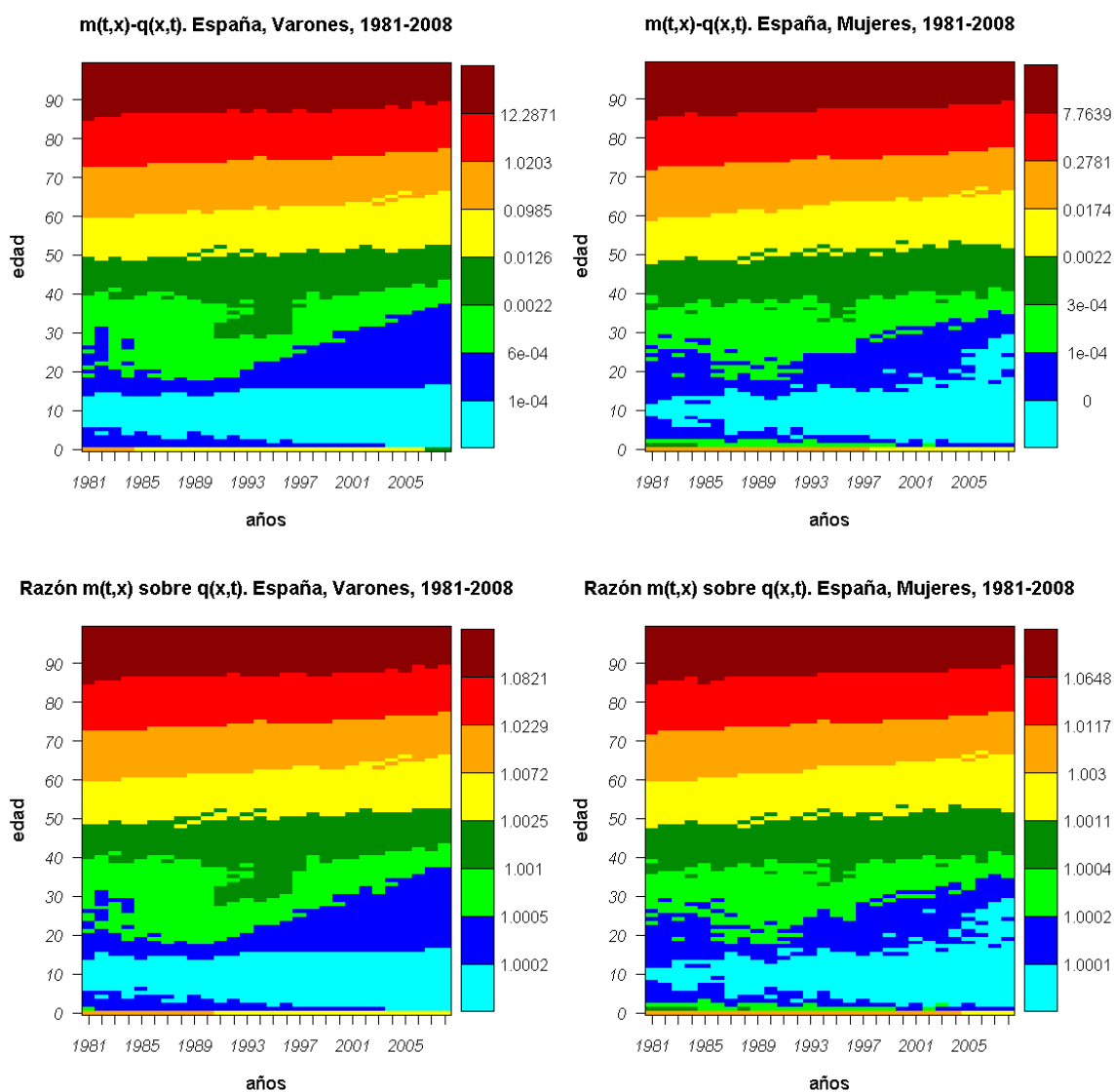


Fuente: Elaboración propia a partir de datos del INE

En la Figura 3.4 se observa que la evolución de estas funciones es bastante regular sin picos ni cambios bruscos, siempre creciente con el tiempo. La esperanza de vida presenta un crecimiento artificial en las últimas edades debido al criterio discrecional de la definición que hace depender estos valores de la tasa. Esta anomalía es más patente en los varones y se corrige con el tiempo y según se baja en la edad al aumentar el tamaño de la población respecto de la que se estima.

En la Figura 3.5 se aprecia que la diferencia entre la tasa y el riesgo es poco significativa, menor cuanto menor es la magnitud de las funciones.

Figura 3.5. Diferencia y razón tasa vs. Riesgo de 1981 a 2008. Hasta 100 años de edad. Izquierda varones, derecha mujeres.



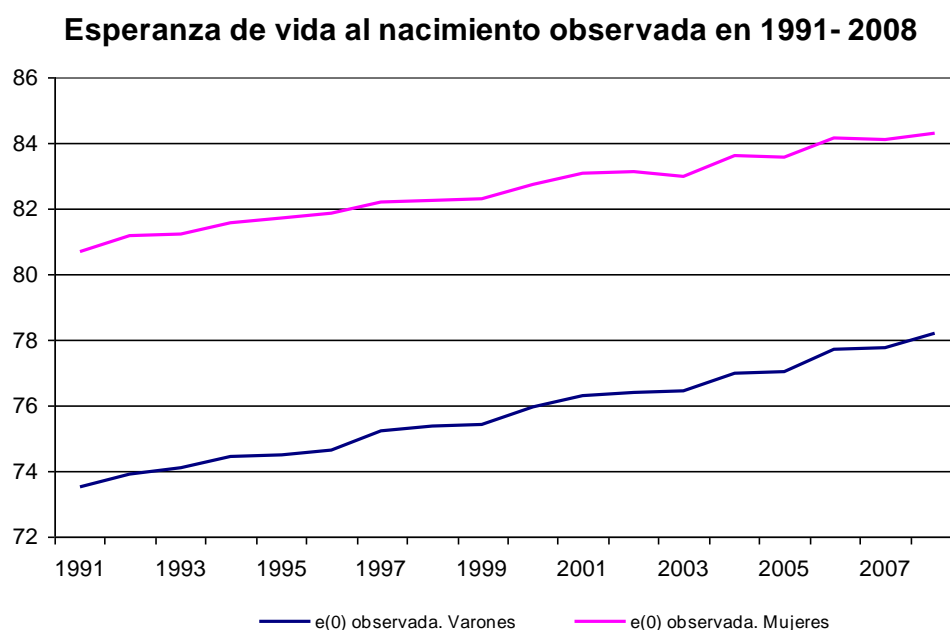
Fuente: Elaboración propia a partir de datos del INE

En la Figura 3.5 se observa que la estructura de las diferencias entre la tasa y el riesgo tiene un comportamiento semejante a las funciones originales. No supera la milésima hasta en torno a los 75 años en varones y para edades superiores en mujeres. En proporción solo excepcionalmente y para edades altas entre las altas son superiores al 10%.

En la Figura 3.6 se puede observar como la esperanza de vida al nacimiento tanto para varones como para mujeres experimenta una tendencia de crecimiento continuo en los 18 años observados. Se observa un crecimiento medio de 0,26 años por año en varones y 0,20 años por año en mujeres. Es decir la población de varones gana más de un año cada 4 y la población de mujeres cada 5.

Sobre la evolución futura, Guijarro y Peláez (2009) proponen un modelo de proyección que explica la historia completa de la esperanza de vida en España.

Figura 3.6. Serie de esperanzas de vida al nacimiento para varones y mujeres. De 1991 a 2008.

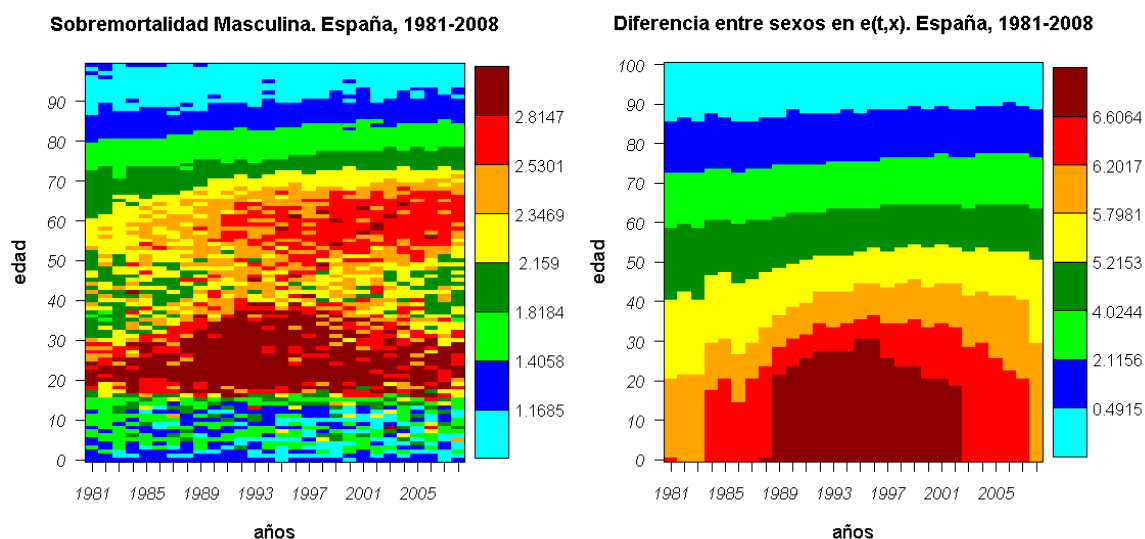


Fuente: Elaboración propia a partir de datos del INE

A continuación se presentan dos indicadores demográficos representativos que ilustran las diferencias entre ambos sexos. El indicador *sobremortalidad masculina* se define como un cociente donde el numerador es el riesgo de mortalidad de los varones

y el denominador es el riesgo de mortalidad de las mujeres. Es mayor de uno excepto para las edades correspondientes a los superlongevos.

Figura 3.7. A la izquierda sobremortalidad masculina. A la derecha diferencia en $e(t,x)$ entre varones y mujeres. De 1981 a 2008. Hasta 100 años de edad.



Fuente: Elaboración propia a partir de datos del INE.

De acuerdo con la Figura 3.7, los valores superiores de *sobremortalidad* masculina se concentran en dos regiones. La principal para edades de los 20 años a casi los 40 crece en los años ochenta y tiene su punto álgido a principios de los noventa donde el riesgo de muerte para los varones triplica al de las mujeres y luego comienza a decrecer paulatinamente. La segunda zona se concentra en las edades entre los 50 y los 70 años y presenta su mayor intensidad en torno al año 2000, donde el riesgo de morir de un varón vuelve a ser de nuevo el triple que el de las mujeres. En las demás edades, y sobre todo en las tempranas y tardías, la mortalidad entre sexos casi se equipara.

De acuerdo con la gráfica de la derecha de la Figura 3.7, la diferencia en esperanza de vida entre mujeres y varones crece para todas las edades a lo largo del tiempo hasta un máximo y luego comienza a decrecer. Para la esperanza de vida al nacimiento el máximo de la diferencia se alcanza a principios de los años noventa para luego caer con rapidez en los últimos años observados. Según se sube en edad el

máximo de la esperanza de vida se va desplazando hacia los últimos años observados. La esperanza de vida a los 90 años alcanza su máximo en el 2005.

La representación de las superficies de mortalidad de las Figuras de este capítulo amplían las posibilidades que dan los clásicos gráficos en dos dimensiones y permiten el estudio demográfico por generación, proporcionando nuevas herramientas y abriendo vías de trabajo para la vista entrenada de demógrafos que sepan ver más allá de los resultados numéricos en bruto y que escapen a las aptitudes de un estudio estadístico como es este trabajo, cuyo objetivo se orienta más hacia el análisis de la mortalidad provincial y la estimación y proyección de la mortalidad de las provincias.

3.3. Tablas de mortalidad de las provincias españolas

Para el cálculo de las tablas de mortalidad provincial se ha utilizado la misma metodología que la empleada para el total nacional de España salvo que se cierran a 95 años, lo que da 96 edades, excepto Ceuta y Melilla, que se cierran a 90 años, lo que da 91 edades. La principal razón para cerrar a edades más tempranas es la poca consistencia de los datos para poblaciones de pequeño tamaño. Puede parecer que son tablas de mortalidad de menor nivel cuando en realidad se trata edades muy avanzadas para el estudio de la mortalidad. Extender la edad a los 100 años es un reto y sobrepasar los 100 años una osadía. Pese a la creciente demanda de información sobre superlongevos los efectivos observados son tan pocos que su estudio supone un desafío a la estadística, la ciencia de los grandes números.

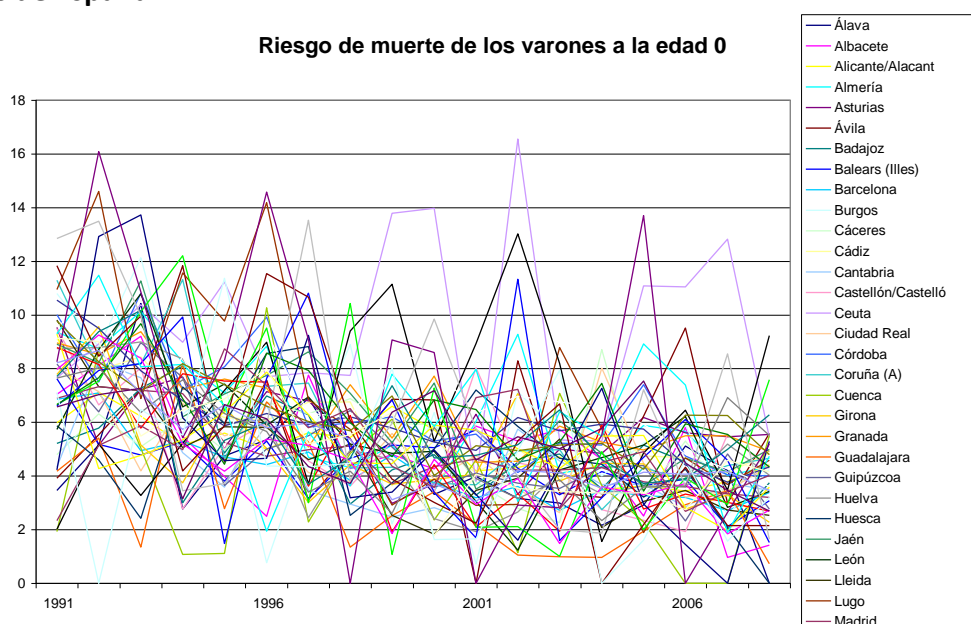
Estas tablas de mortalidad son las que utiliza el INE como paso intermedio en su metodología de 2009 para calcular las agregadas que son las que publica.

En este apartado se expone un resumen de todos los resultados obtenidos al calcular las tablas de mortalidad de las provincias, que son un total de 1.434.816 registros.

3.3.1. El riesgo de muerte

Una primera aproximación al estudio de la mortalidad en las provincias se acomete con el análisis del riesgo de muerte en los varones a la edad 0. En la Figura 3.8 se observa una nube de valores que, en conjunto, siguen una tendencia aunque sin mostrar patrones individuales de comportamiento. Se aprecian fluctuaciones irregulares, un proceso errático, quizá hasta aleatorio en torno al valor del total nacional. A lo mejor un ruido blanco, i.e., un proceso estocástico con una media y una varianza que dependen, no de las provincias, sino única y exclusivamente del año t y la edad x , en este caso la edad 0.

Figura 3.8. Serie de riesgos de mortalidad de los varones a la edad 0 para las 52 provincias de España.



Fuente: Elaboración propia a partir de datos del INE

La representación gráfica conjunta de las 52 provincias es confusa. Como aproximación, y para facilitar la comprensión del fenómeno demográfico de la mortalidad en cada provincia, se ha seleccionado un conjunto de provincias representativo atendiendo al tamaño poblacional y la distribución geográfica.

En la tabla 3.1 se puede observar la desigualdad en la distribución de la población por provincias. Las dos de mayor tamaño tienen el 25% de la población total. En particular la mayor es 89,83 veces más grande que la más pequeña.

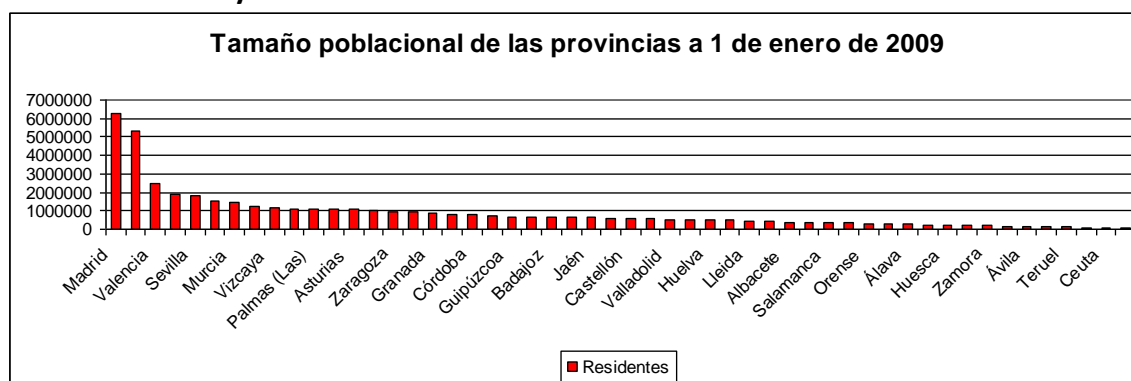
Tabla 3.1: Población de España y cada una de sus provincias a 1 de enero de 2009.

| | | | | | | | |
|------------------|----------|------------------------|---------|-------------|--------|-------------|--------|
| Total Nacional | 45828172 | | | | | | |
| Madrid | 6295011 | Santa Cruz de Tenerife | 1004397 | Castellón | 591667 | Rioja (La) | 315718 |
| Barcelona | 5345603 | Zaragoza | 945600 | Cantabria | 576418 | Álava | 307656 |
| Valencia | 2515246 | Pontevedra | 942890 | Valladolid | 523431 | Guadalajara | 236399 |
| Alicante | 1884876 | Granada | 902524 | Ciudad Real | 518227 | Huesca | 222315 |
| Sevilla | 1849858 | Tarragona | 792097 | Huelva | 502689 | Cuenca | 214693 |
| Málaga | 1562606 | Córdoba | 786326 | León | 483714 | Zamora | 193902 |
| Murcia | 1443383 | Gerona | 725166 | Lérida | 427426 | Palencia | 170724 |
| Cádiz | 1211219 | Guipúzcoa | 691751 | Cáceres | 405790 | Ávila | 169891 |
| Vizcaya | 1136654 | Almería | 679776 | Albacete | 396764 | Segovia | 161835 |
| Coruña (La) | 1123033 | Badajoz | 674649 | Burgos | 366102 | Teruel | 145820 |
| Palmas (Las) | 1072188 | Toledo | 656564 | Salamanca | 347787 | Soria | 93159 |
| Baleares (Islas) | 1070066 | Jaén | 655469 | Lugo | 346020 | Ceuta | 72561 |
| Asturias | 1058923 | Navarra | 614526 | Orense | 326987 | Melilla | 70076 |

Fuente: Elaboración propia a partir de datos del INE

La media de la distribución es 881.311 residentes por provincia y entre las 9 primeras provincias reúnen la mitad de la población.

Figura 3.9. Diagrama de barras con la población de las provincias de España ordenadas de mayor a menor.



Fuente: Elaboración propia a partir de datos del INE

Las provincias en verde de la tabla 3.1 son las seis provincias tomadas como muestra representativa para estudiar la evolución del riesgo de muerte provincial.

La más pequeña es Lérida, que tiene un tamaño suficiente como para que la serie de sus datos no fluctúe en exceso. Por debajo de este tamaño las oscilaciones del riesgo y la tasa en algunos casos son violentas de un año a otro. Para descartar una posible influencia territorial se seleccionan provincias repartidas por todo el territorio

nacional desde el Norte, Guipúzcoa, hasta el Sur, Cádiz; de Oeste, La Coruña, a Este, Lérida, pasando por las provincias interiores de Badajoz y Valladolid para descartar posibles influencias como puede ser tener costa.

En las gráficas de la Figura 3.10 se presentan los riesgos observados en estas seis provincias y el total para España a lo largo de 18 años para distintas edades.

Figura 3.10. Serie de riesgos de mortalidad de los varones a las edades 0, 20, 40, 60 y 80 en las provincias de Badajoz, Cádiz, La Coruña, Lérida, Guipúzcoa, Valladolid y en total nacional de España. A la izquierda varones, a la derecha mujeres.

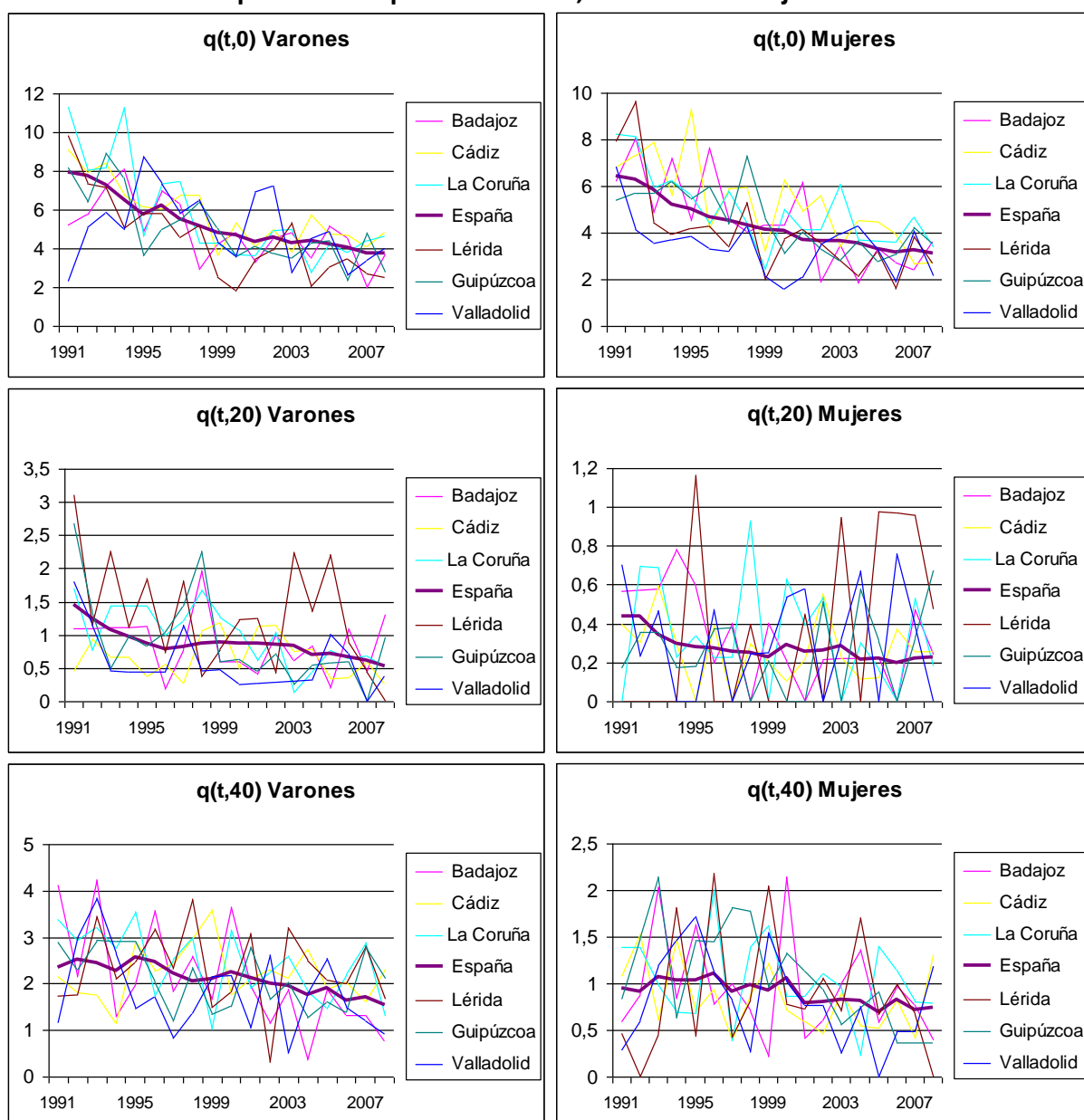
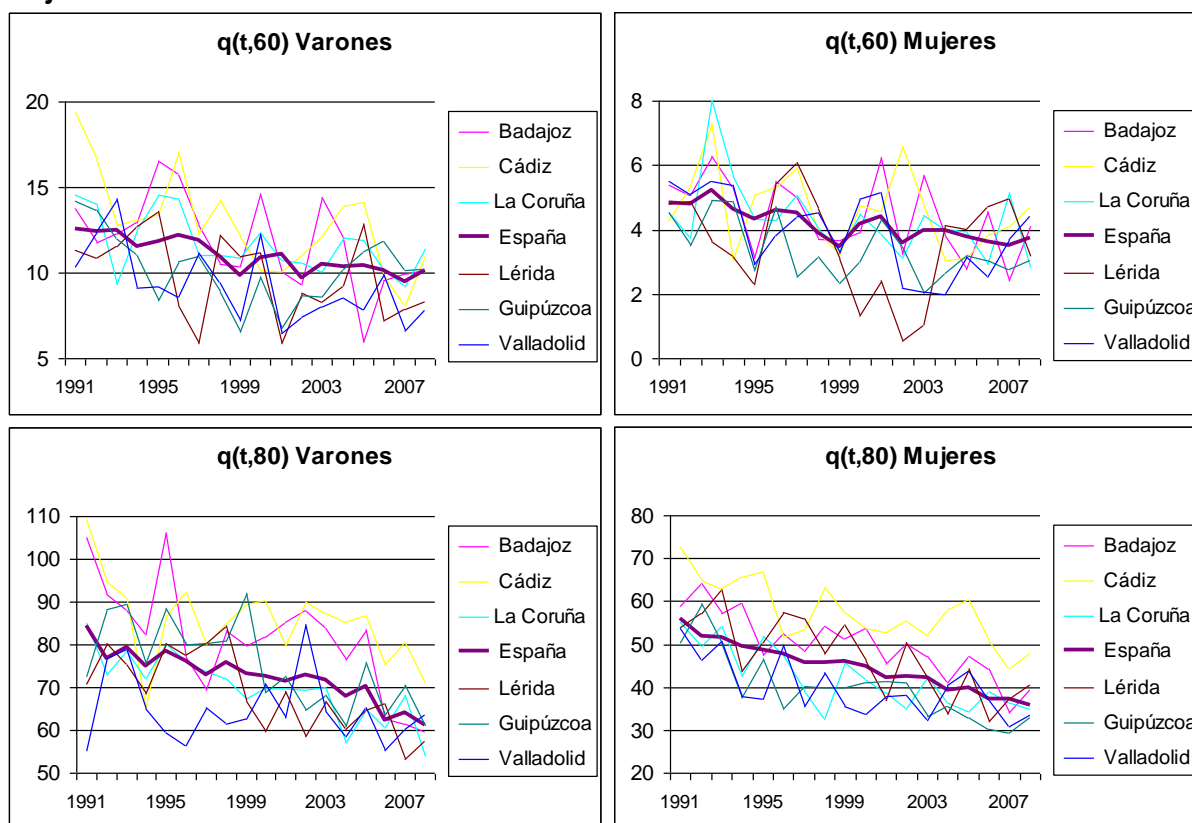


Figura 3.10 (Cont.). Serie de riesgos de mortalidad de los varones a las edades 0, 20, 40, 60 y 80 en las provincias de Badajoz, Cádiz, La Coruña, Lérida, Guipúzcoa, Valladolid y en total nacional de España. A la izquierda varones, a la derecha mujeres.



Fuente: Elaboración propia a partir de datos del INE

Es interesante observar en la Figura 3.10 que el riesgo de muerte de las provincias se distribuye en torno al valor del riesgo de muerte para España y, más aún, que la dispersión de los valores provinciales crece con el valor del riesgo de muerte. Es decir la variabilidad depende de forma creciente del riesgo de muerte³.

Para constatar estas dos apreciaciones se analiza a continuación el caso particular de las 5 edades y para las 6 provincias representadas en las gráficas de la Figura 3.10.

Se toman las edades 0, 20, 40, 60 y 80 como muestra representativa que facilita la comprensión del análisis. Naturalmente el mínimo absoluto de la función riesgo de muerte se obtiene a edades en torno a 10 años de edad como se puede observar en las gráficas de $q(t,x)$ de la Figura 3.2.

³ Es importante observar que cada gráfica tiene distinta escala. Todas están en milésimas y van de 1,2 para las mujeres a los 20 años hasta 110 en los varones a los 80 años.

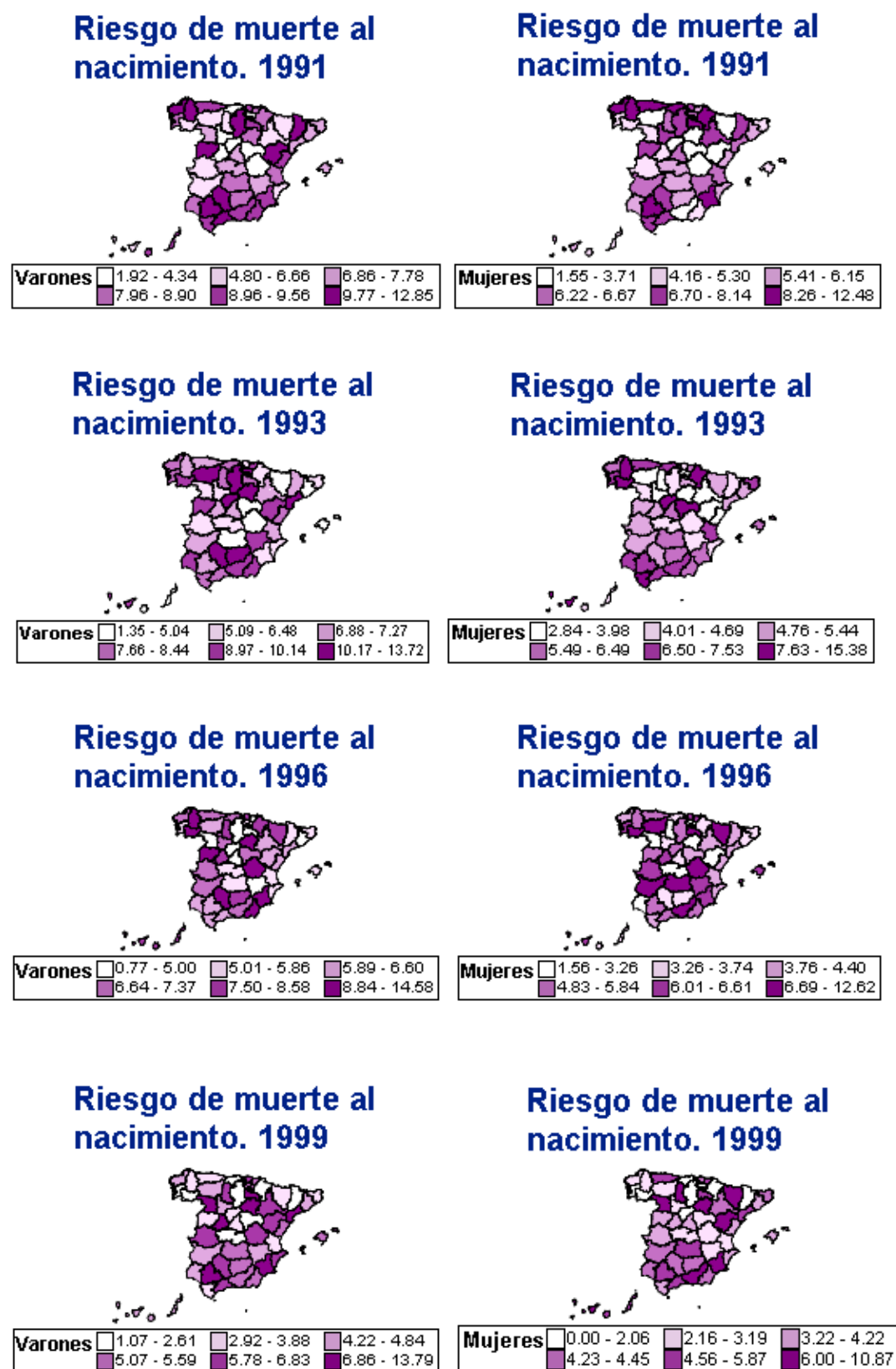
Para nuestro estudio particular utilizamos las edades de 0, 20, 40, 60 y 80 años. En primer lugar, respecto a los varones, el menor valor del riesgo de muerte en España se tiene a los 20 años de edad. Con un valor en torno a una milésima se tiene un rango de la distribución para los riesgos de las provincias de unas 3 milésimas. Le sigue la edad de 40 años a la que para un valor observado de 2 milésimas se tiene un rango en torno a unas 4 milésimas. Luego va la edad de 0 años con un valor de en torno a 6 milésimas y un rango de unas 8 milésimas. A continuación los 60 años de edad. Un valor en torno a las 12 milésimas y un rango de unas 10 milésimas. Finalmente la edad de 80 años. Con un valor en torno a 70 milésimas y un rango de unas 30 milésimas.

Para mujeres el menor valor del riesgo de muerte en España es a los 20 años de edad. Para un valor en torno a 0,3 milésimas se tiene un rango de la distribución para los riesgos de las provincias de una milésima. Le sigue la edad de 40 años a la que para un valor observado de una milésima se tiene un rango en torno a unas 2 milésimas. Luego va la edad de 60 años con un valor de en torno a 5 milésimas y un rango de unas 6 milésimas. A continuación los 0 años de edad. Un valor en torno a las 6 milésimas y un rango de unas 8 milésimas. Finalmente la edad de 80 años. Con un valor en torno a 55 milésimas y un rango de unas 25 milésimas.

Cabe preguntarse si existen patrones provinciales. Si la distribución geográfica la mortalidad presenta cierta continuidad. Los mapas de distribución provincial del riesgo de muerte muestran ciertas provincias más propensas a caer en la parte superior o en la parte inferior del rango del riesgo de muerte provincial. A pesar de la tendencia de mantenerse en la franja superior o inferior del rango todos los años observados hay alguna provincia que parece haberse desmarcado de su patrón. Cuando parece que se ha determinado una tendencia para una provincia ésta presenta un dato fuera de la tónica.

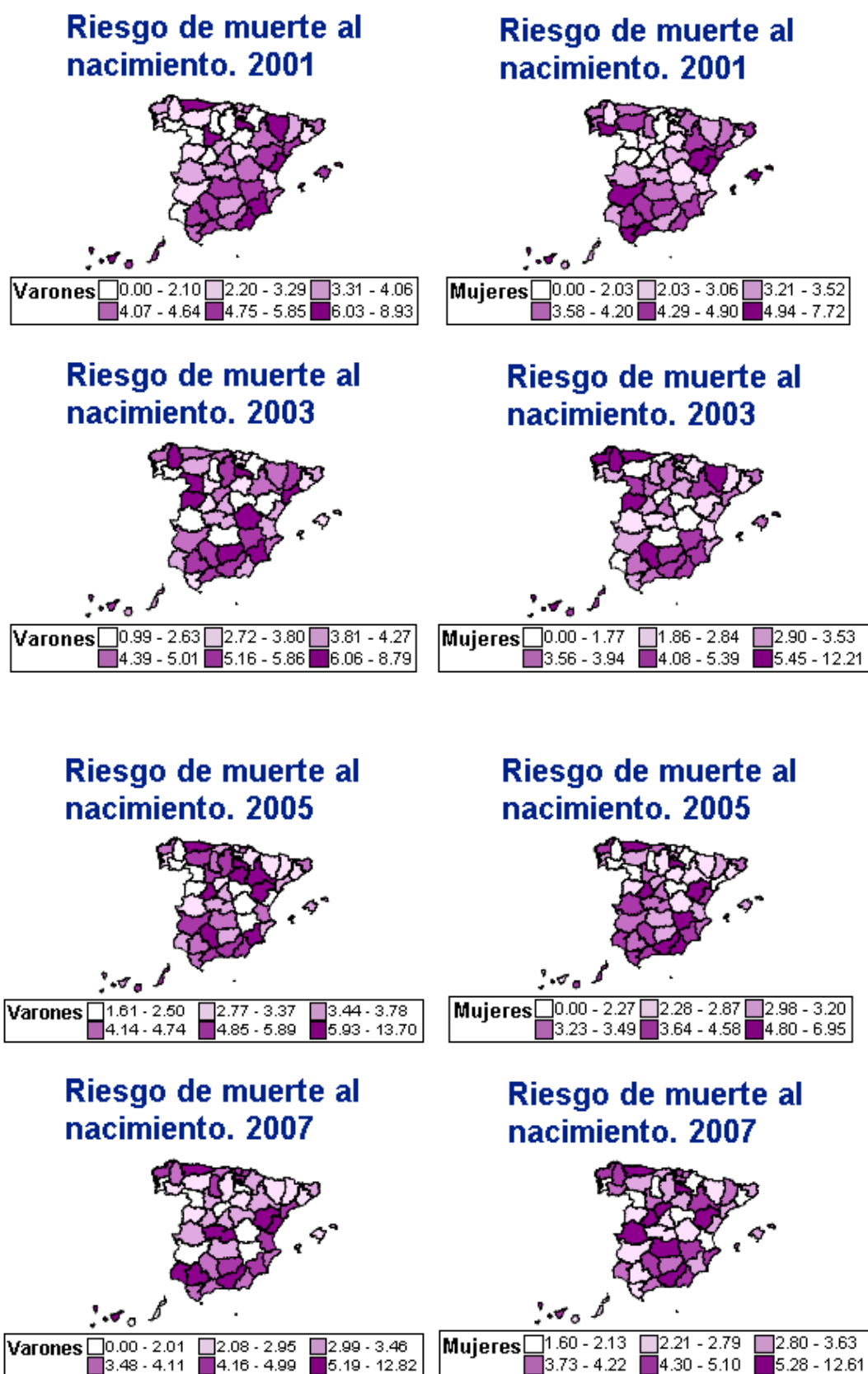
De acuerdo con la Figura 3.11 y la Figura 3.12 se observa que la distribución territorial del nivel del riesgo de muerte de las provincias cada año no sigue ningún patrón lógico. La provincia de León pasa de tener una mortalidad en los varones respecto a las demás favorable en 1991 a pasar a desfavorable en 1993 para luego estabilizar en los años 1995, 1997 y volver a subir y bajar en 2005 y 2007.

Figura 3.11. $q(t,0)$ para todas las provincias los años 1991, 1993, 1996 y 1999.



Fuente: Elaboración propia a partir de datos del INE.

Figura 3.12. $q(t,0)$ para las provincias en los años 2001, 2003, 2005 y 2007.



Fuente: Elaboración propia a partir de datos del INE.

Aunque el riesgo de muerte de las provincias se distribuye en torno al valor del riesgo de muerte para España no se puede establecer ninguna relación determinista entre el valor del riesgo para España y el valor del riesgo de una provincia. Lo que sí se puede intuir es que provincias con tamaño poblacional grande tienen un comportamiento más estable que las provincias de tamaño poblacional pequeño. Aún así no se observa en ninguna provincia un patrón sistemático que se repita a lo largo del tiempo. El comportamiento del riesgo en las provincias se presta poco a ser modelizado. Es coyuntural y poco predecible.

3.3.2. La esperanza de vida

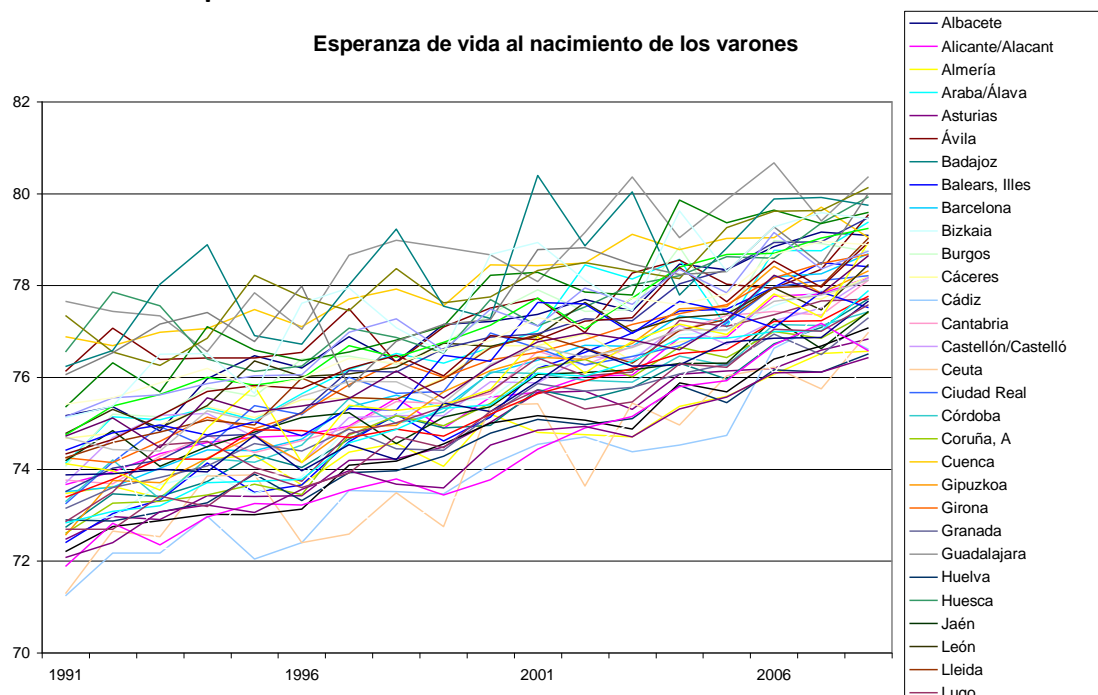
Frente a los inconvenientes ya comentados del riesgo de muerte, la esperanza de vida presenta importantes ventajas. En primer lugar se trata de una variable más conocida y comprensible.

En particular la esperanza de vida al nacimiento es el indicador más utilizado y que resume mejor el comportamiento de la mortalidad en una población. Aunque es cierto que, como todo resumen, encubre información y puede llevar a interpretaciones confusas.

Por construcción, la esperanza de vida a la edad x depende de la esperanza de vida a edades superiores. Por tanto, a diferencia del riesgo de muerte, una variable excesivamente coyuntural y expuesta a sucesos raros, la esperanza de vida es más consistente al depender de una combinación de muchos sucesos de probabilidad pequeña, germen de las distribuciones normales. Esta característica se acentúa cuanto menor es la edad con una pérdida paulatina de consistencia según aumenta la edad y siendo una de las causas principales que fuerza al cierre de la tabla de mortalidad a menor edad.

En segundo lugar utilizar la esperanza de vida elimina el factor cero en el soporte. Para que la esperanza de vida a la edad x sea cero los supervivientes tienen que fallecer instantáneamente en el comienzo de año. Esta propiedad es una ventaja de la variable riesgo que sí contempla el cero como parte del soporte. Se trata de una

Figura 3.13. Serie de riesgos de mortalidad de los varones a la edad 0 para las 52 provincias de España.

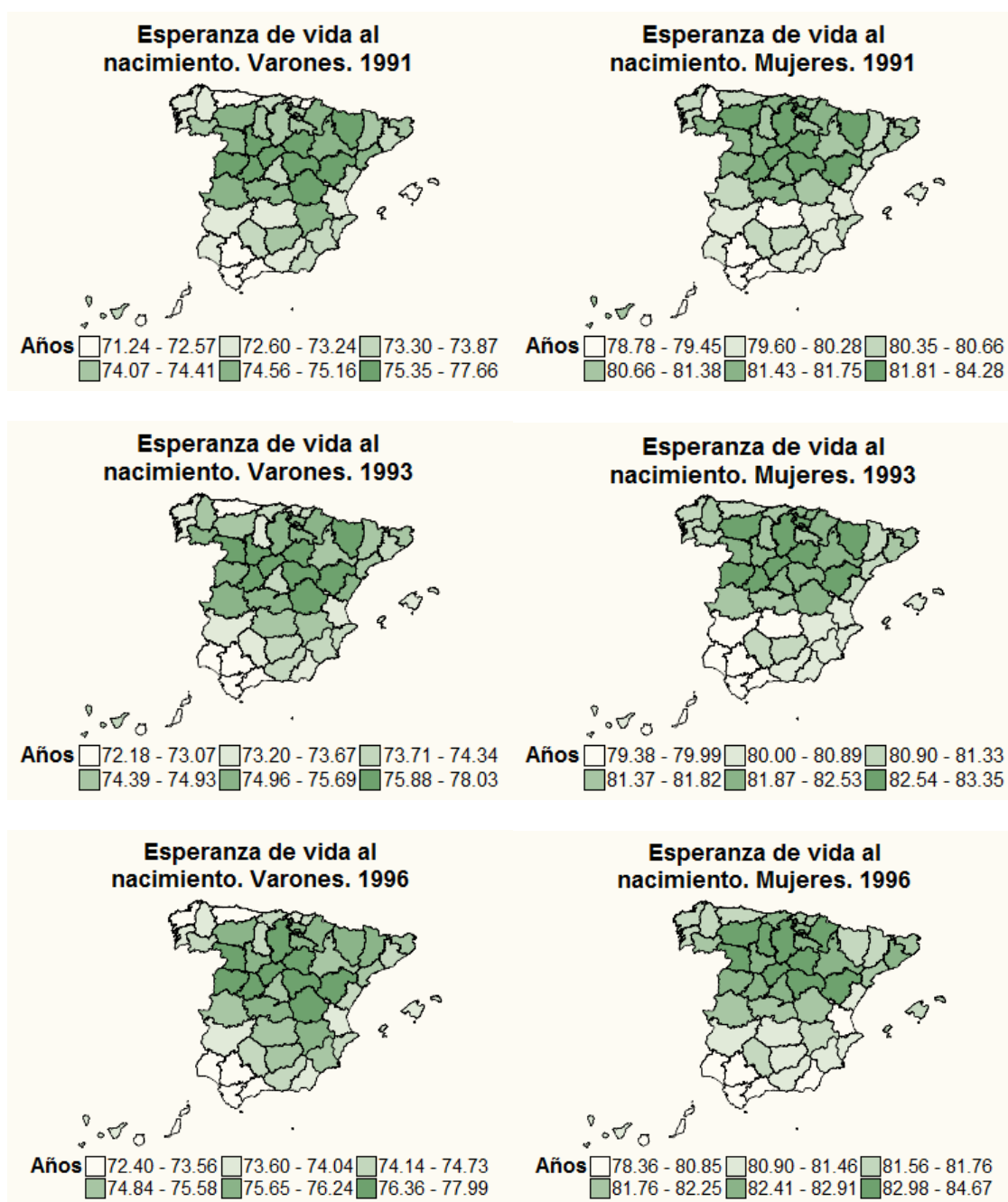


Fuente: Elaboración propia a partir de datos del INE

En la serie de mapas de coropletas de la Figura 3.14, la Figura 3.15 y la Figura 3.16 se observa que la distribución geográfica de la esperanza de vida al nacimiento, tanto para varones como para mujeres, describe una superficie con contigüidad entre provincias sin saltos excesivamente bruscos. El siguiente paso evolutivo es la continuidad de la superficie. Esta fase abre un interesante conjunto de líneas de investigación. Desde la medición a los criterios de elección de la muestra y como colofón una superficie continua que aporta una novedosa visión de la mano de una innovadora forma de entender el concepto de esperanza de vida.

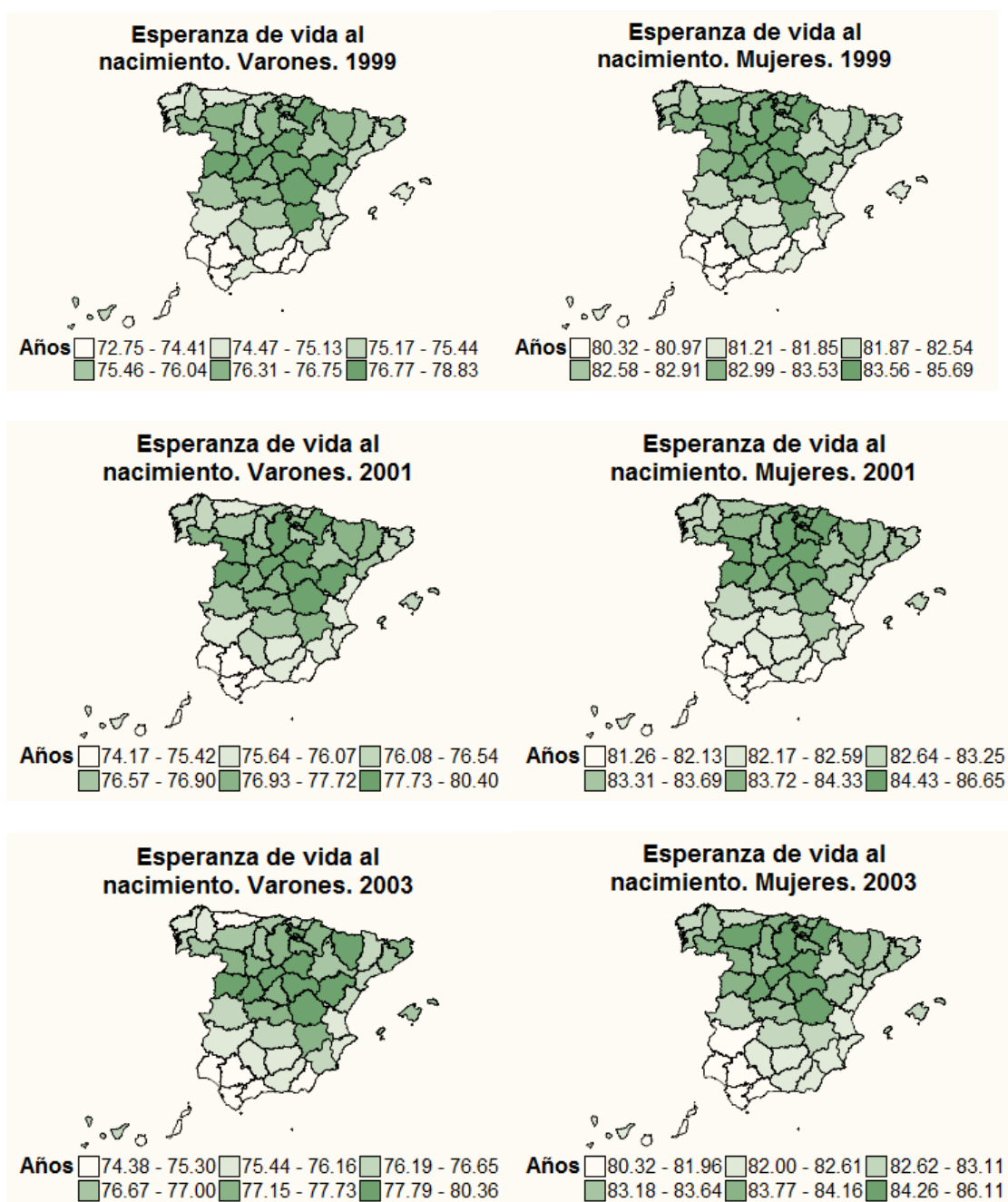
Las provincias costeras presentan una esperanza de vida al nacimiento inferior a las provincias interiores. Además la esperanza de vida crece según se va de sur a norte y de oeste a este. Estas posiciones relativas se conservan a lo largo de los casi veinte años con ligeros cambios. La componente sur va tomando más peso y la región con menor esperanza de vida se desplaza por la costa sur hacia el este a medida que pasamos al siglo XXI.

Figura 3.14. Esperanza de vida al nacimiento en las provincias los años 1991, 1993 y 1996. A la izquierda varones, a la derecha mujeres.



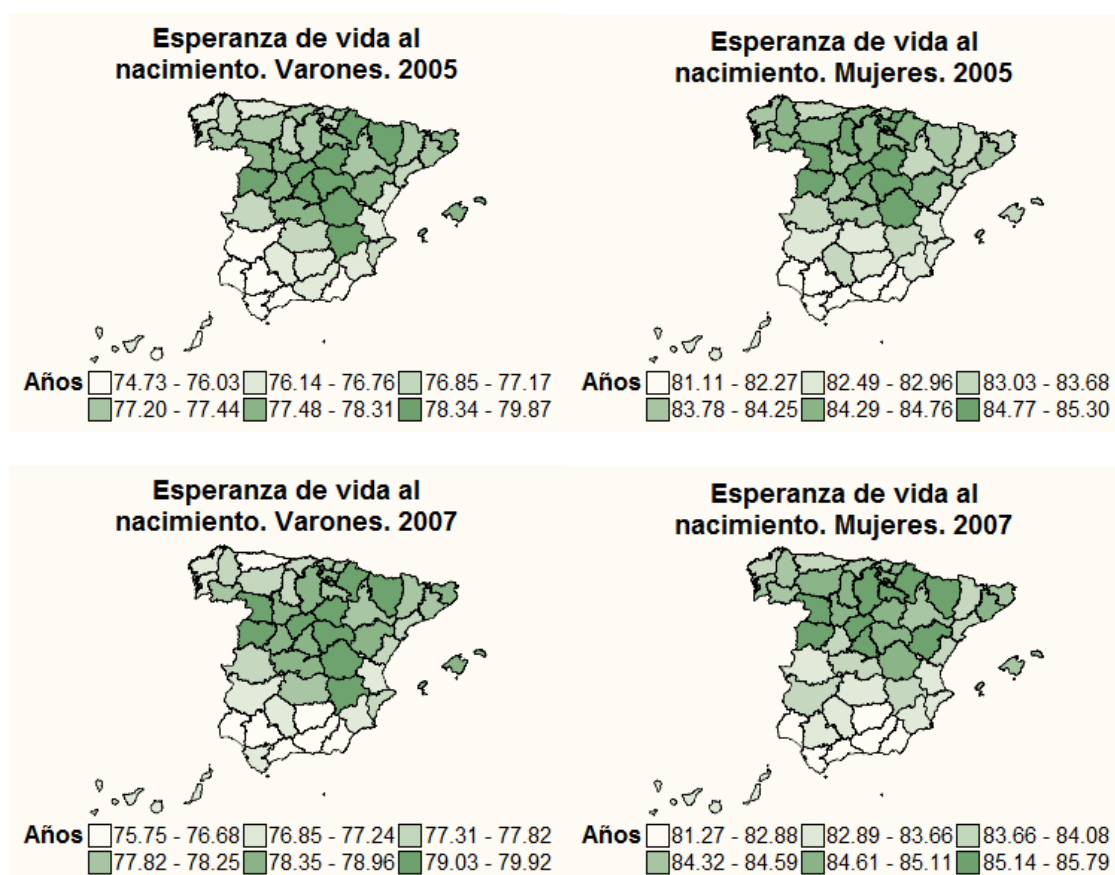
Fuente: Elaboración propia a partir de datos del INE

Figura 3.15. Esperanza de vida al nacimiento en las provincias los años 1999, 2001 y 2003. A la izquierda varones, a la derecha mujeres.



Fuente: Elaboración propia a partir de datos del INE

Figura 3.16. Esperanza de vida al nacimiento en las provincias los años 2005 y 2007. A la izquierda varones, a la derecha mujeres.

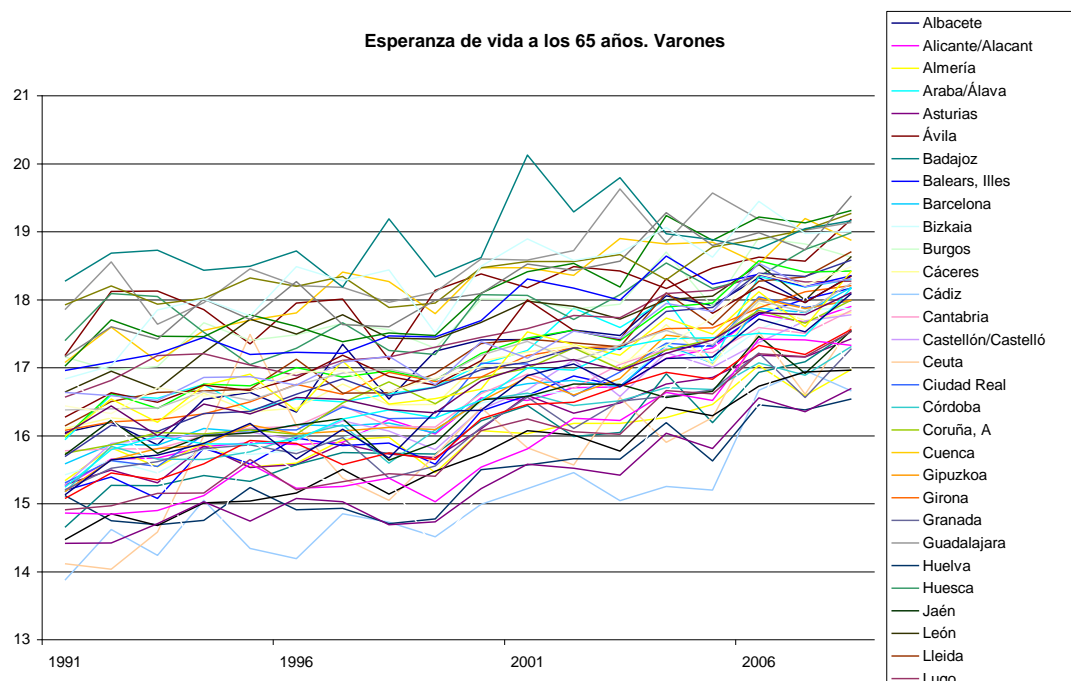


Fuente: Elaboración propia a partir de datos del INE

La esperanza de vida al nacimiento es interesante como indicador y primera aproximación unidimensional para la clasificación de las provincias de acuerdo a su mortalidad, pero su uso de forma aislada puede llevar a conclusiones erróneas. Dos esperanzas de vida iguales pueden encubrir una distribución en la incidencia de la mortalidad por edades completamente diferente. La esperanza de vida al nacimiento es un resumen y deja de tener en cuenta gran cantidad de información correspondiente al resto de esperanzas de vida a las distintas edades.

Por ejemplo, como se puede ver en la Tabla 3.2, la esperanza de vida al nacimiento de los varones en Ciudad Real el año 1997 y en Gerona el año 1999 coincide en 76 años. Sin embargo, de acuerdo con la tabla 3.4 la esperanza de vida calculada a la edad de 65 es 16,42 años en Ciudad Real en 1997 y de 16,79 años en Gerona en 1999.

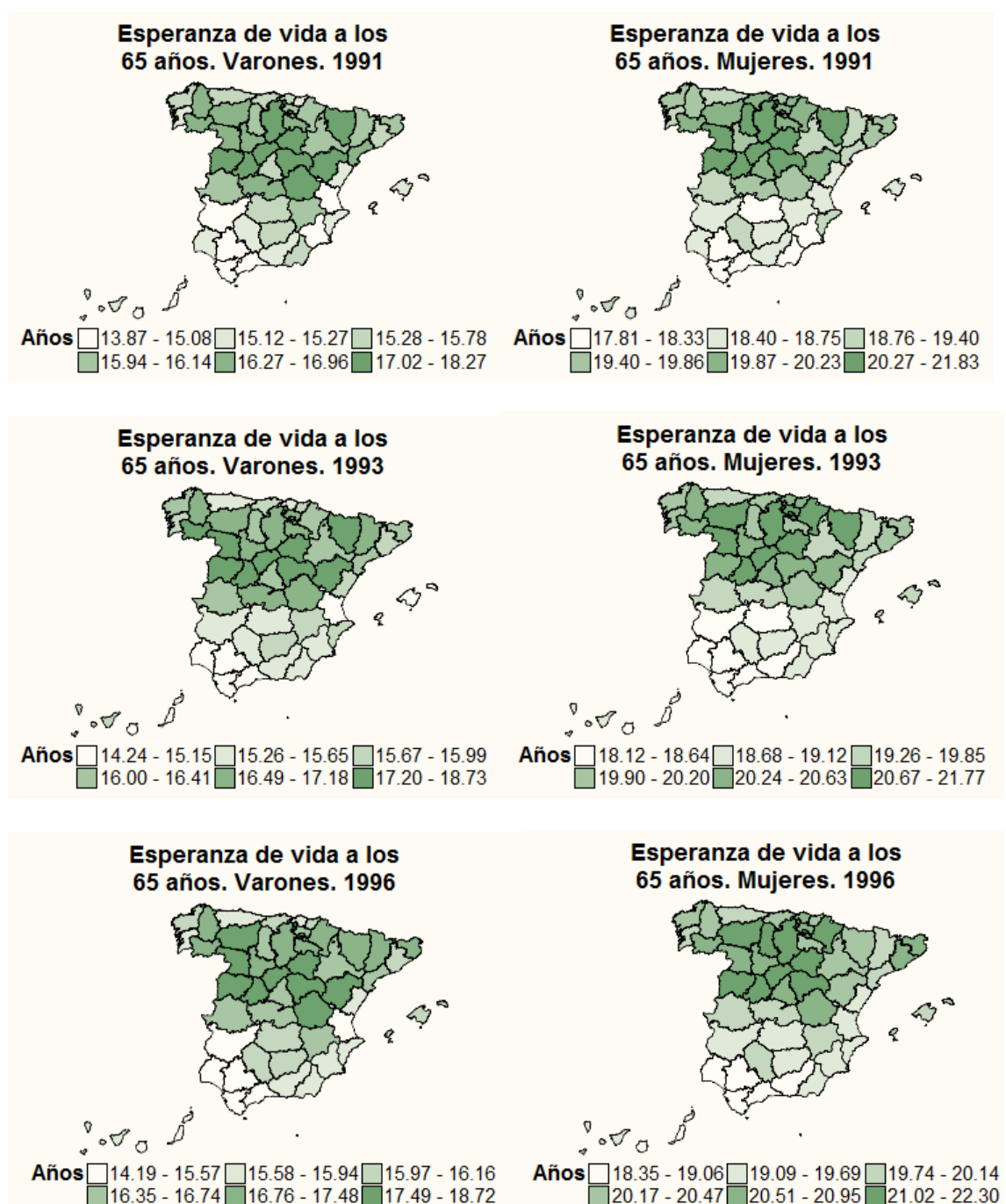
Figura 3.17. Series de tiempo de esperanza de vida a los 65 años en las provincias desde 1991 hasta 2008. Varones



Fuente: Elaboración propia a partir de datos del INE

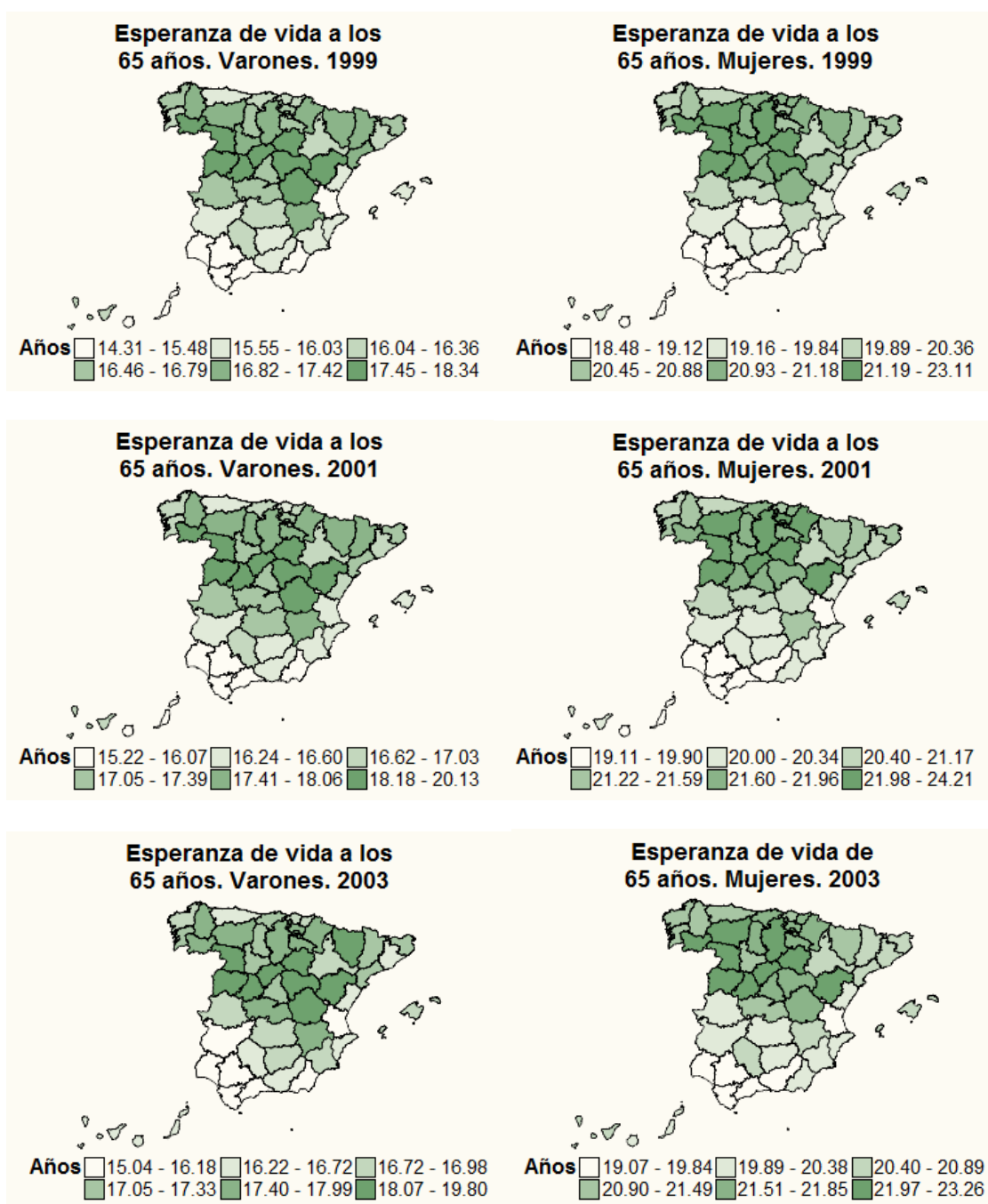
La distribución geográfica de la esperanza de vida a los 65 años que se ve en la Figura 3.18 es similar a la observada para la esperanza de vida al nacimiento. La superficie es relativamente continua y las regiones con valores superiores son interiores y se sitúan en la zona norte. Las provincias con valores inferiores se sitúan en las regiones costeras y septentrionales.

Figura 3.18. Esperanza de vida a los 65 años en las provincias los años 1991, 1993 y 1996. A la izquierda varones, a la derecha mujeres.



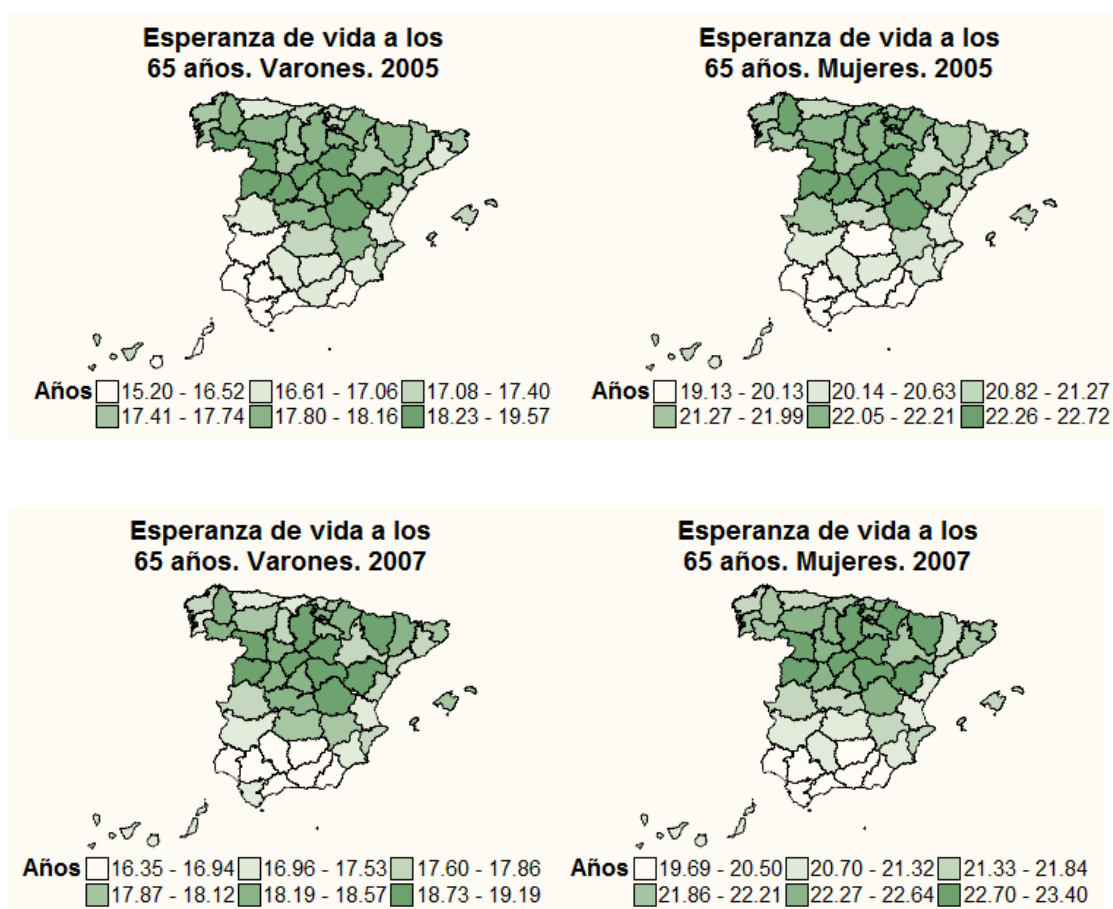
Fuente: Elaboración propia a partir de datos del INE

Figura 3.19. Esperanza de vida a los 65 años en las provincias los años 1999, 2001 y 2003. A la izquierda varones, a la derecha mujeres.



Fuente: Elaboración propia a partir de datos del INE

Figura 3.20. Esperanza de vida a los 65 años en las provincias los años 2005 y 2007. A la izquierda varones, a la derecha mujeres.



Fuente: Elaboración propia a partir de datos del INE

Por cada esperanza de vida a cada edad que se incluye en el conjunto de datos se extiende el espacio con una nueva dimensión al espacio de las esperanzas de vida que describe la mortalidad. La máxima dimensión de espacio para analizar la mortalidad de todas la provincias es 91. Una por cada una de las 91 edades para las que se calcula la esperanza de vida para todas las provincias. Si se incluyen los ejes que corresponden a la esperanza de vida desde los 91 años a los 95 años, Ceuta y Melilla no aparecen porque no tienen coordenada.

3.4. Clasificación de las provincias por esperanza de vida

El análisis de datos ofrece varias alternativas para el estudio del espacio de las esperanzas de vida. Distintas razones dan lugar a elegir un análisis de componentes principales de un espacio con las 19 dimensiones correspondientes a las edades de 0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 y 85 años.

La razón para terminar a la edad de 85 años es la excesiva exposición de la esperanza de vida a la edad de cierre de la tabla y próximas a factores coyunturales que ocasionan grandes fluctuaciones y falta de consistencia debido al pequeño tamaño de la muestra a partir de la cual se estima el valor. La influencia de los sucesos raros se amortigua a medida que baja la edad y la diferencia entre Ceuta y Melilla que cierran a 90 y el resto de provincias que cierra a 95 se atenúa. Algunos análisis sugieren la posibilidad de que cerrar a 85 sigue incorporando demasiados factores coyunturales. Una línea de investigación abierta es bajar la edad por debajo de 70 años. Bajo esta elección está el conflicto entre la pérdida de información de edades especialmente interesantes en el rango de los longevos y la consistencia deseable en la estimación de edades, excesivamente expuestas a factores coyunturales.

El motivo para utilizar únicamente las esperanzas de vida de la tabla de mortalidad por grupos quinquenales se debe a que el resto no aporta información adicional relevante. La esperanza de vida se construye recursivamente, lo que implica que existe una gran dependencia grande entre las distintas variables o ejes que representan las esperanzas de vida a edades próximas.

Dentro de las alternativas que ofrece la minería de datos esta el análisis cluster. En nuestro caso, efectuar un análisis cluster obliga a tomar una serie de decisiones arbitrarias. Hay que establecer un criterio de agrupación que dará lugar a conjuntos de agrupaciones de provincias diferentes cada año. Es difícil establecer patrones de comportamiento por grupos y la identificación de provincias representantes.

El análisis de componentes principales, en cambio, proporciona un resumen de la información a partir de las primeras componentes que se pueden interpretar de

forma sencilla. Además da opción a definir unas componentes comunes a todos los años que permiten estudiar la evolución a lo largo del tiempo de las provincias.

En primer lugar, se reduce, mediante un Análisis de Componentes Principales (ACP), la información que contienen las 19 variables consideradas en el análisis a lo largo de los 18 años que comprende el período de análisis considerado. En todos los años del estudio, se comprueba que, al seleccionar dos componentes principales, se retiene una varianza de al menos el 97%⁴, como se muestra en la Tabla 3.6.

Tabla 3.6. Porcentaje de Varianza explicada por las dos primeras Componentes Principales en el ACP sobre las esperanzas de vida a las distintas edades.

| Año | Porcentaje de Varianza Explicada. Esperanza de vida de Varones | Porcentaje de Varianza Explicada. Esperanza de vida de Mujeres |
|------|---|---|
| 1991 | 97,828 | 98,542 |
| 1992 | 97,593 | 97,567 |
| 1993 | 98,223 | 98,852 |
| 1994 | 97,657 | 98,788 |
| 1995 | 97,494 | 98,651 |
| 1996 | 97,240 | 98,382 |
| 1997 | 97,419 | 98,961 |
| 1998 | 98,178 | 99,048 |
| 1999 | 97,618 | 99,166 |
| 2000 | 97,855 | 98,960 |
| 2001 | 98,136 | 99,098 |
| 2002 | 98,129 | 98,907 |
| 2003 | 97,915 | 99,329 |
| 2004 | 97,795 | 99,188 |
| 2005 | 97,525 | 99,194 |
| 2006 | 97,160 | 98,773 |
| 2007 | 97,065 | 98,997 |
| 2008 | 97,842 | 98,957 |

Fuente: Elaboración propia a partir de datos del INE

Este análisis preliminar permite concluir que el ACP es un método óptimo para resumir la información que contienen los indicadores de esperanza de vida a las distintas edades consideradas por género, en cada uno de los años considerados. Este método permite hacer un estudio transversal en cada año analizado, pero no permite establecer comparaciones longitudinales combinando y resumiendo toda la

⁴ En cualquier caso, el porcentaje de varianza que acumula la primera componente siempre es superior al 85%, lo que la configura como una componente de tamaño. Esto es muy habitual en los estudios en los que las variables que se pretenden resumir mediante ACP están altamente correlacionadas, como es el caso que se analiza en este trabajo.

información en un único plano, puesto que las dos primeras componentes principales son específicas para cada año. En el apartado 3.4.1 se va a presentar el análisis transversal de los datos a través del Análisis de Componentes Principales aplicado sobre la información de la esperanza de vida por grupos quinquenales durante los años 1991 a 2008. En el apartado 3.4.2 se aplica una técnica que permite realizar un estudio longitudinal de la información de las variables analizadas.

3.4.1. Análisis por año

La primera componente principal en cada año es una media de todas las variables en la que las esperanzas de vida a las edades superiores tienen una ponderación ligeramente inferior. Esta componente explica globalmente la esperanza de vida con lo que es un indicador directo de la mortalidad y por tanto está estrechamente relacionada con la esperanza de vida al nacimiento. La segunda componente principal es una combinación lineal de las variables que da mayor peso a las edades más avanzadas, explica la esperanza de vida en el rango superior de edades.

De acuerdo con el plano que forman las dos componentes principales se delimitan las regiones correspondientes a los cuatro cuadrantes que se definen eligiendo como centro de coordenadas la posición que ocupa el caso del Total Nacional. Los cuadrantes primero y cuarto presentan una mortalidad más favorable en general que el Total Nacional. En los cuadrantes primero y segundo la mortalidad es más favorable en las edades avanzadas.

En la Figura 3.21 aparecen los planos de las dos primeras componentes principales para cada año, en donde cada caso se sitúa expresado por rangos para facilitar la comprensión del mismo.

Para tratar de clasificar las provincias en grupos, se va a comparar su situación relativa en cada periodo con la situación del Total Nacional. Por la configuración del ACP realizado aquellas provincias que se sitúen por encima del Total Nacional en ambas componentes tendrán un mejor comportamiento en términos de mayor esperanza de vida.

Figura 3.21. Plano de las dos primeras componentes principales por año. Varones.

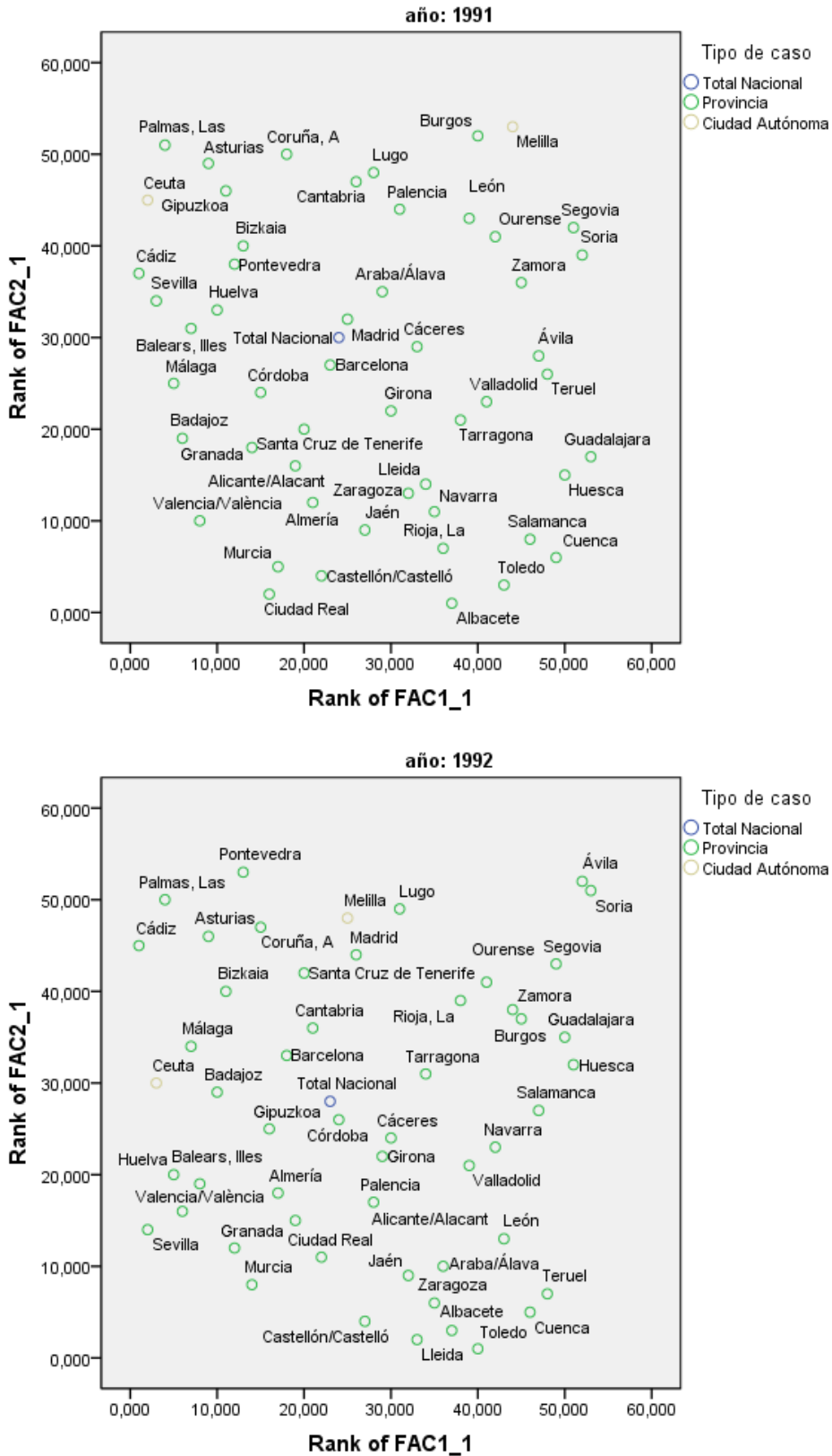


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

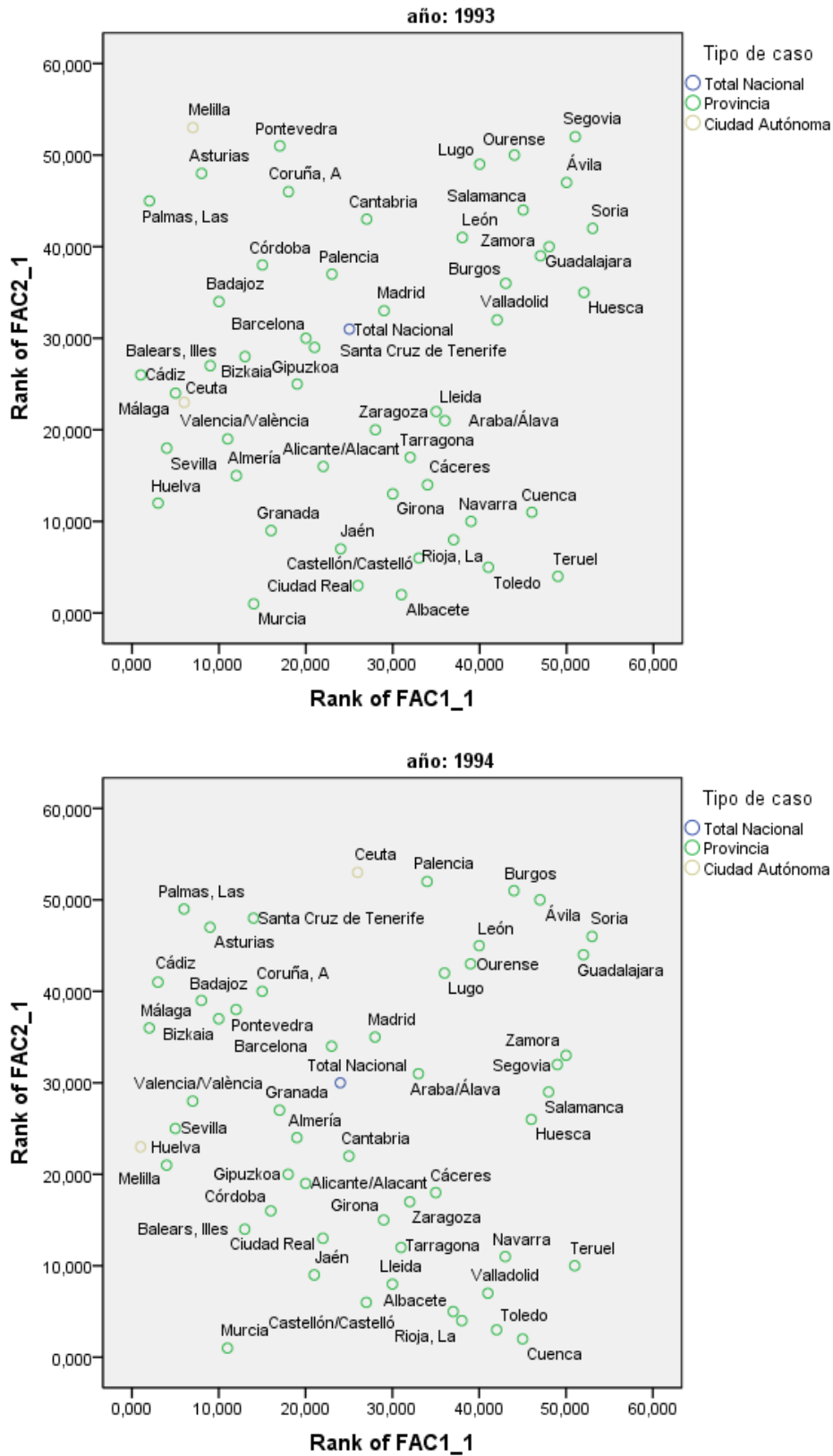


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

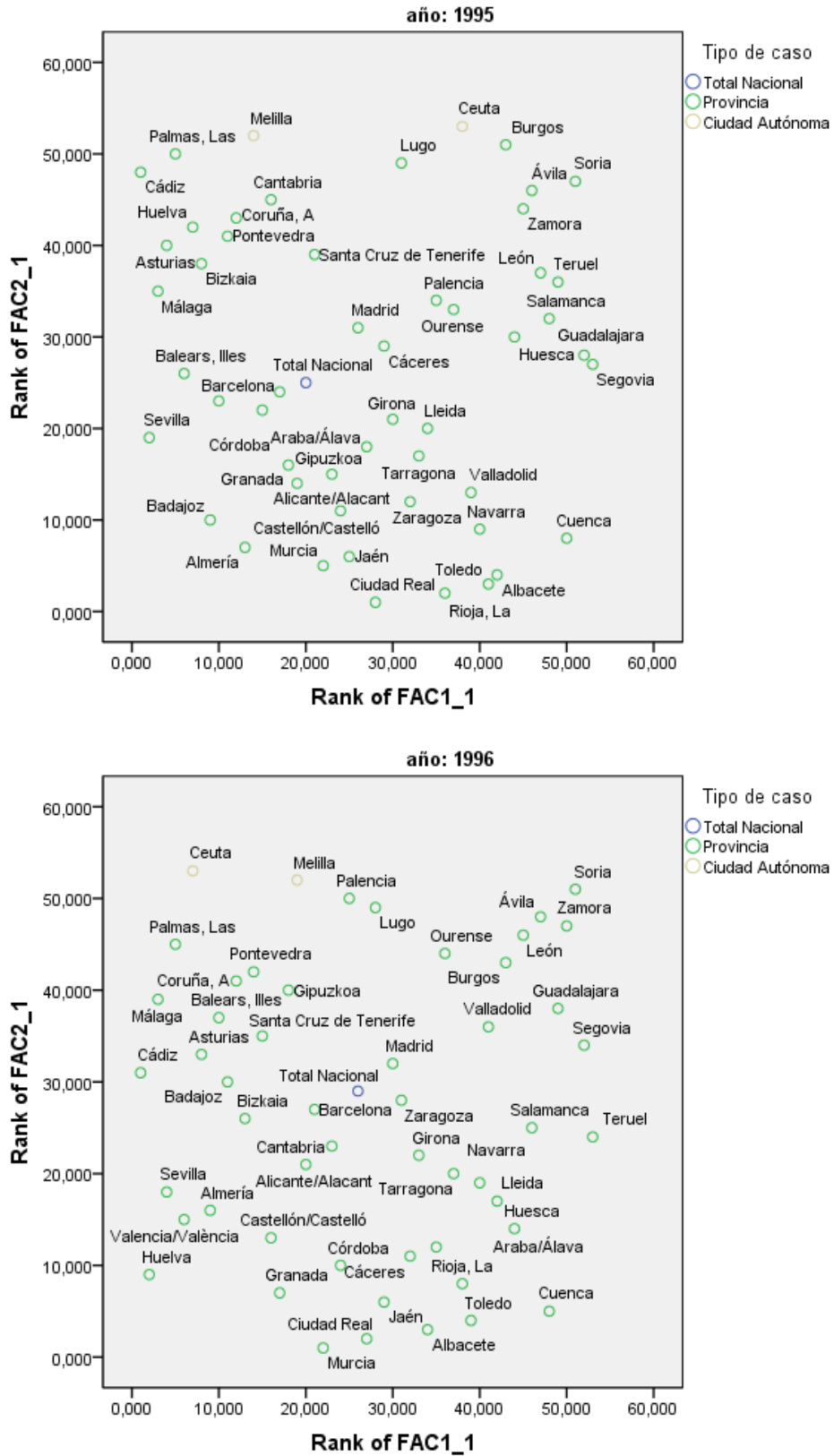


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

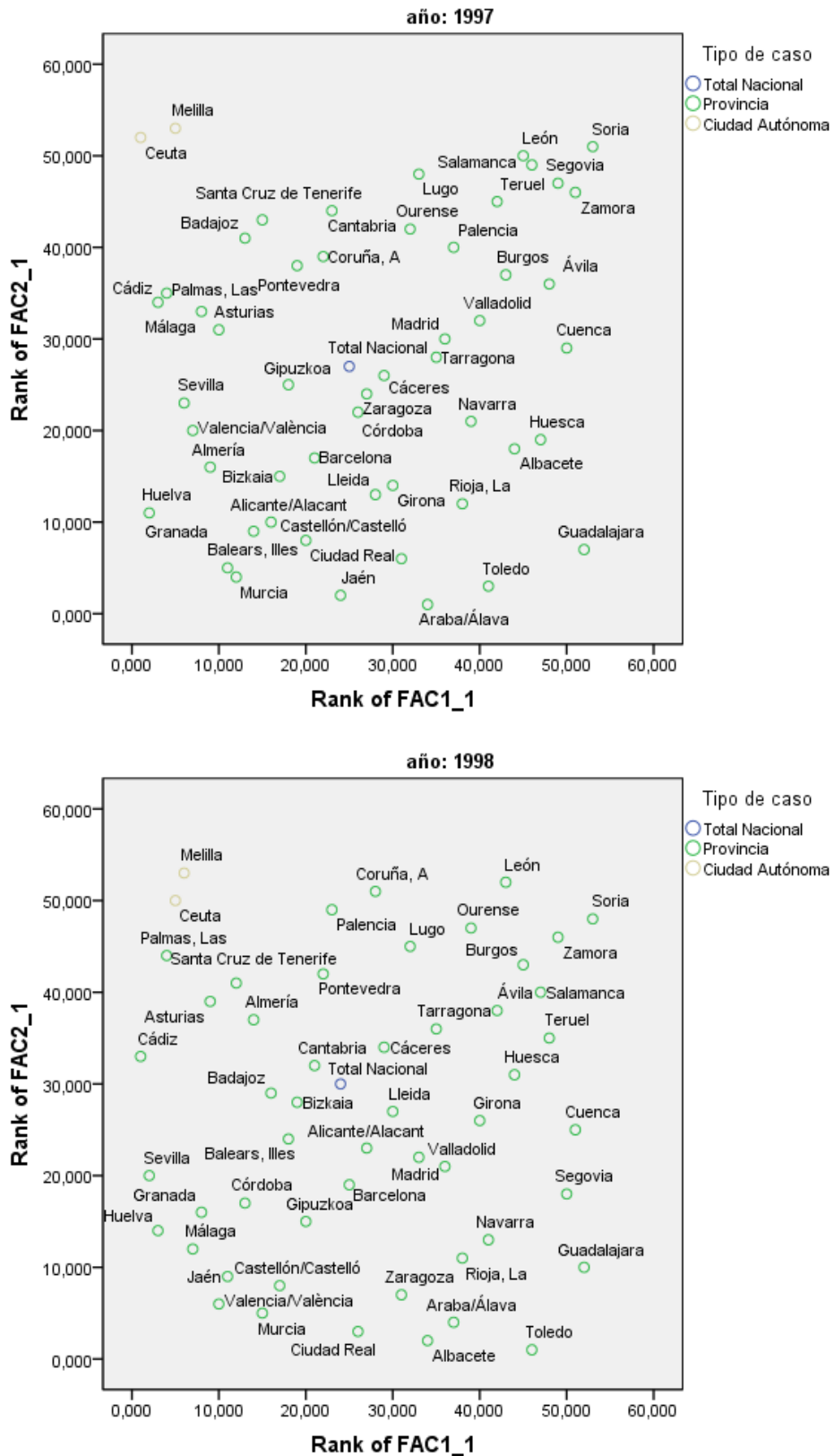


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

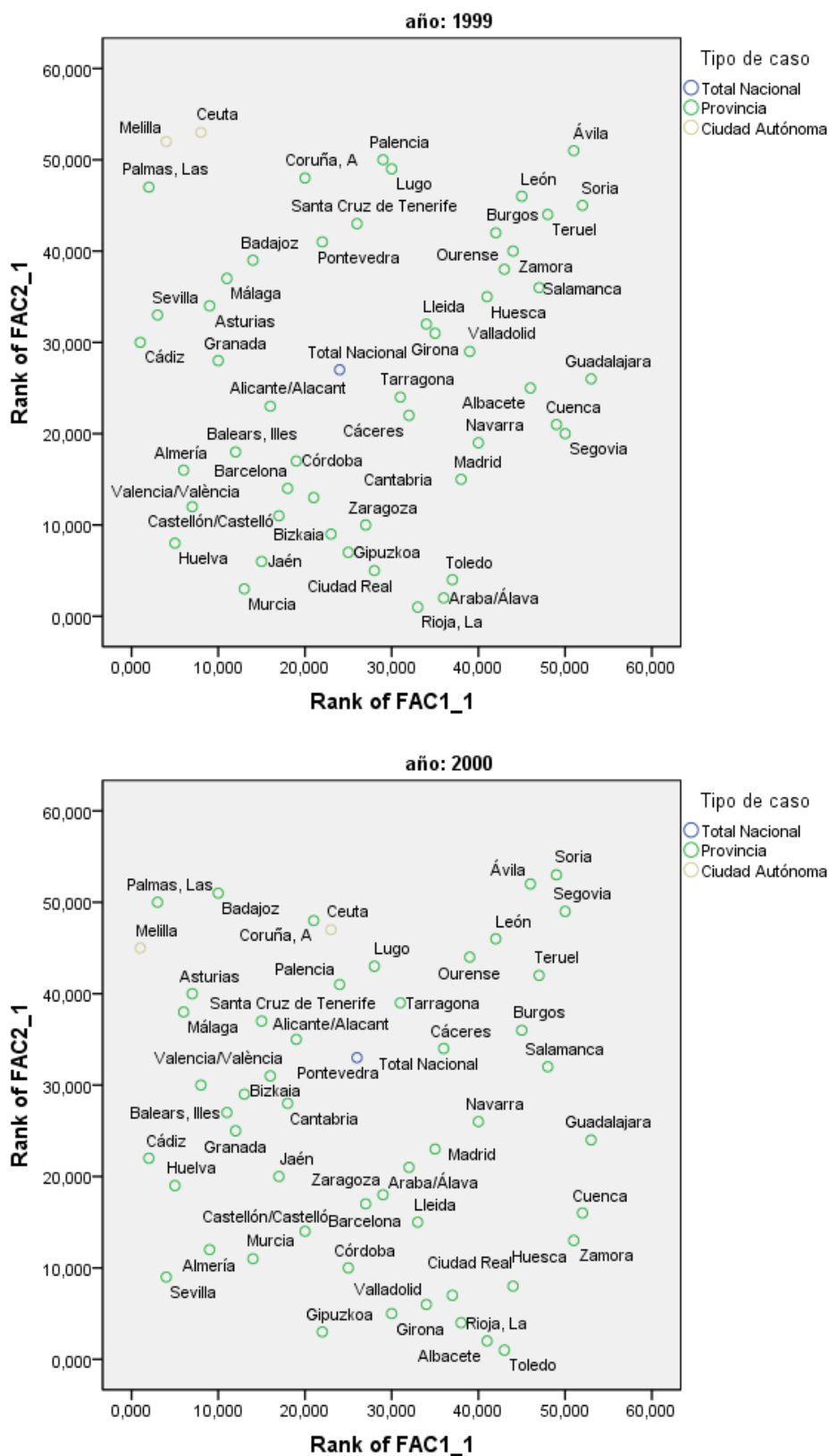


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

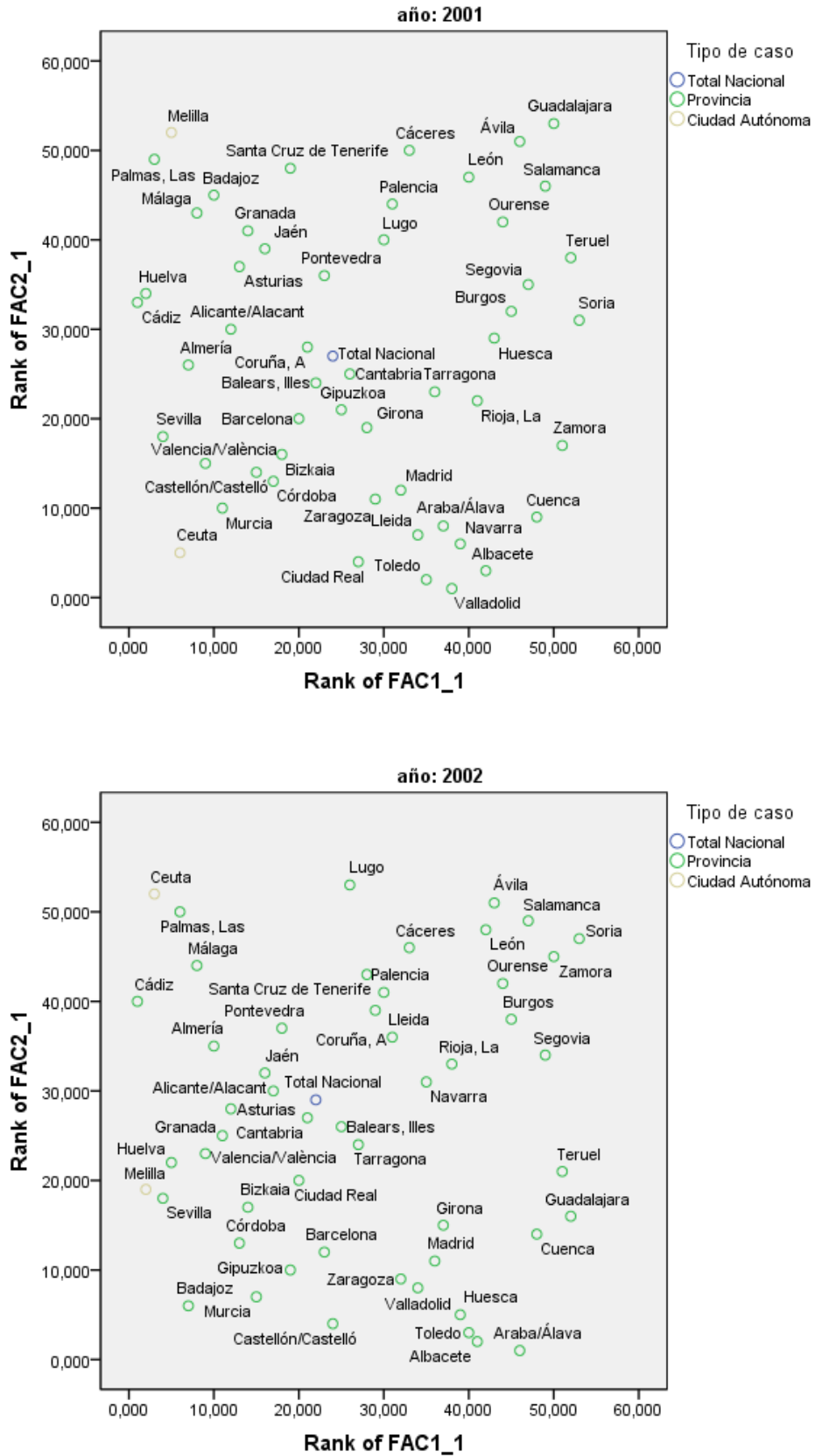


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

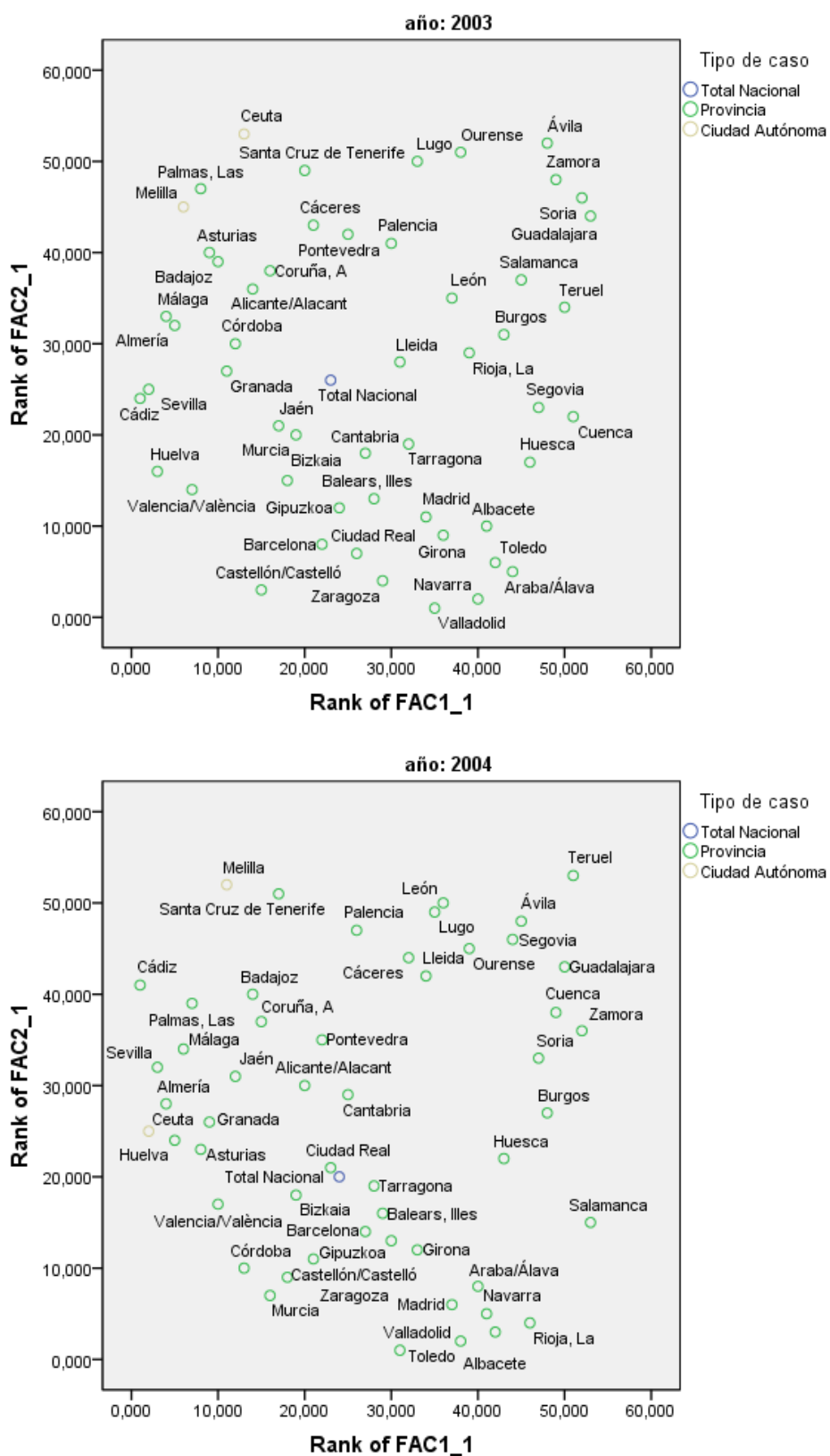


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.

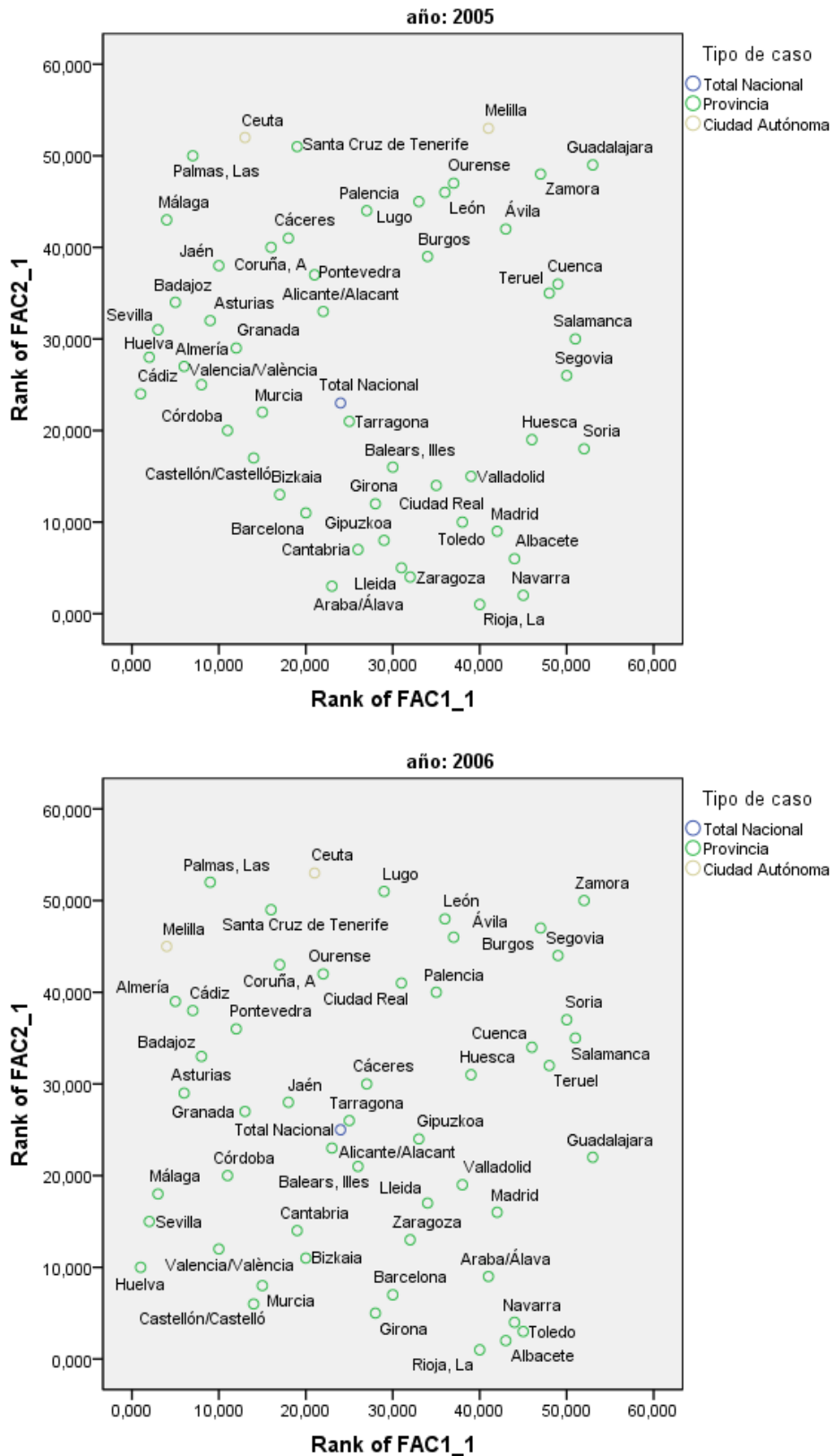
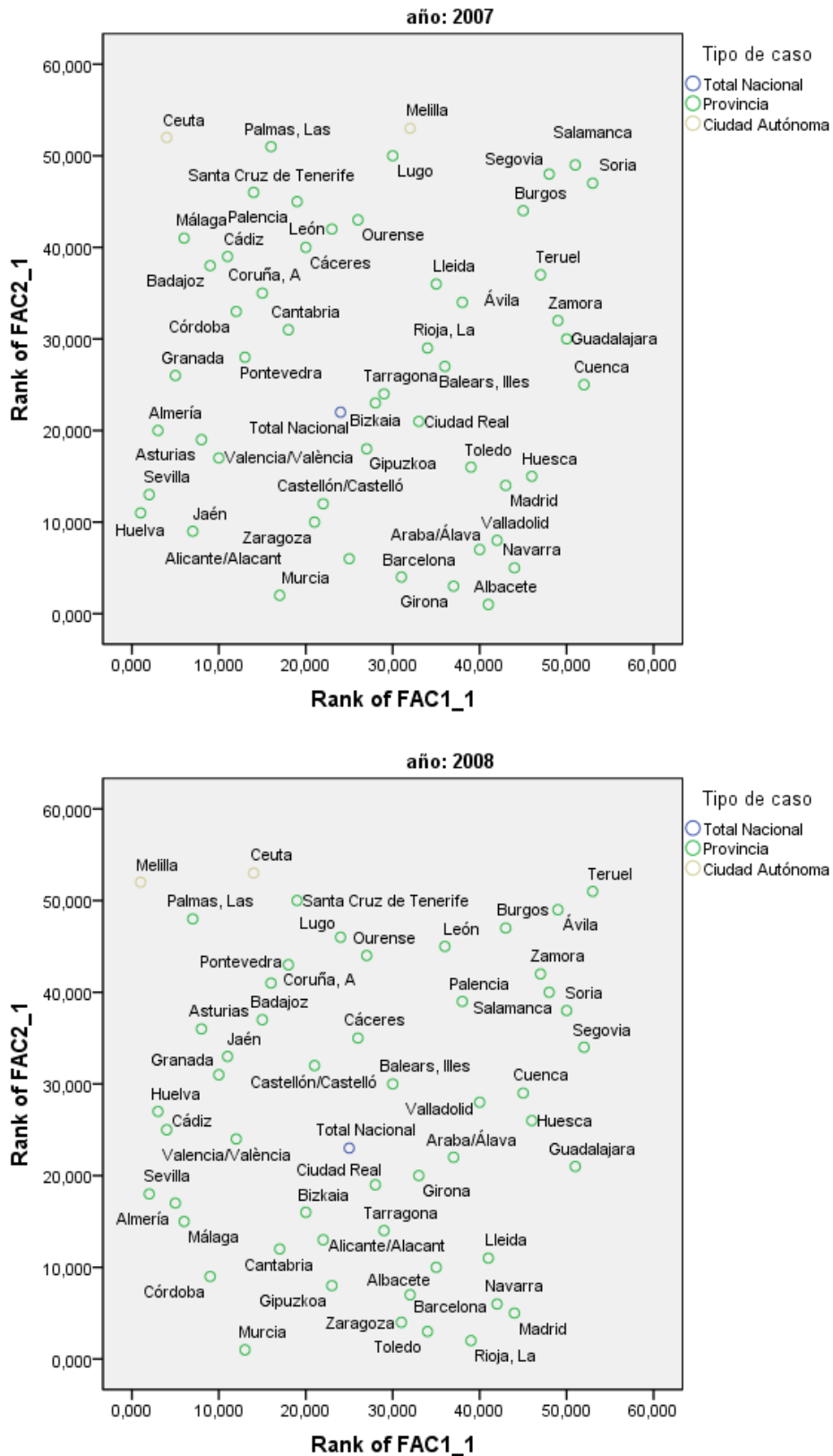


Figura 3.21 (cont.). Plano de las dos primeras componentes principales por año. Varones.



Estudio de la Mortalidad en España y sus Provincias

Atendiendo a la clasificación por cuadrantes, se agrupan las provincias en la Tabla 3.7. Las 4 columnas de la derecha en la Tabla 3.7 proporcionan la frecuencia absoluta de veces que cada provincia aparece en cada cuadrante.

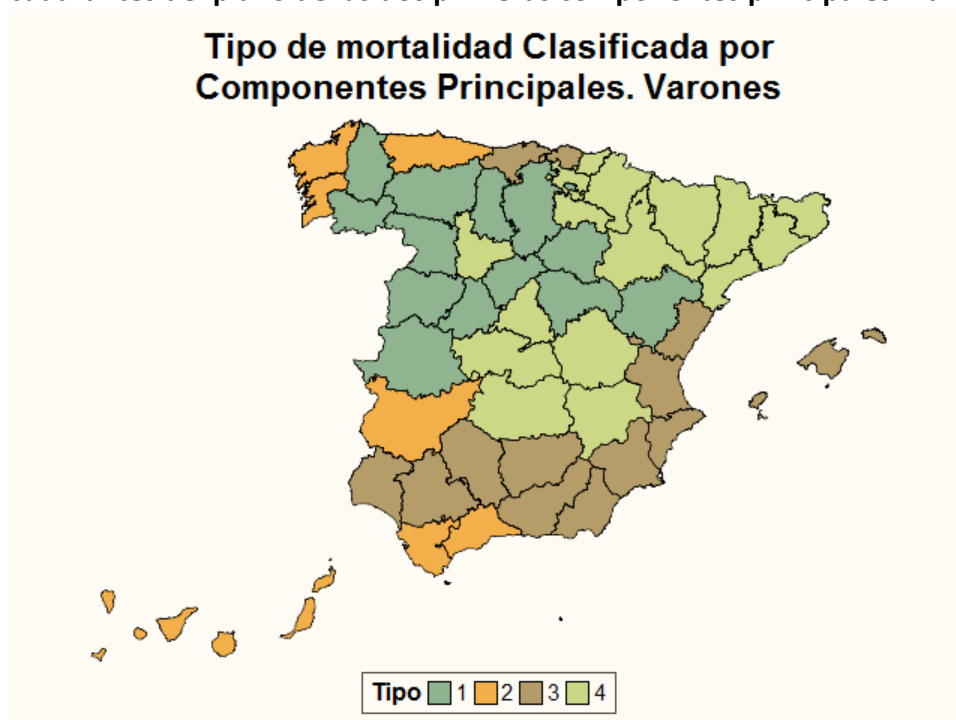
Tabla 3.7. Clasificación de las provincias por cuadrante por año y frecuencias absolutas. Varones.

| | año | | | | | | | | | | | | | | | | | | Número de | | | | |
|------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----------|----|----|----|----|
| | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 1 | 2 | 3 | 4 | |
| Albacete | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 18 |
| Alicante/Alacant | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 4 | 3 | 0 | 6 | 10 | 2 |
| Almería | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 0 | 6 | 12 | 0 | |
| Araba/Álava | 1 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 2 | 0 | 1 | 15 | |
| Asturias | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 0 | 16 | 2 | 0 | |
| Ávila | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 17 | 0 | 0 | 1 | |
| Badajoz | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 14 | 4 | 0 | |
| Balears, Illes | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 1 | 1 | 2 | 3 | 8 | 5 | |
| Barcelona | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 0 | 2 | 9 | 7 | |
| Bizkaia | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 1 | 4 | 13 | 0 | |
| Burgos | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 18 | 0 | 0 | 0 | |
| Cáceres | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 8 | 3 | 0 | 7 | |
| Cádiz | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 15 | 3 | 0 | |
| Cantabria | 1 | 2 | 1 | 4 | 2 | 3 | 2 | 2 | 3 | 3 | 4 | 3 | 4 | 1 | 4 | 3 | 2 | 3 | 3 | 5 | 6 | 4 | |
| Castellón/Castelló | 3 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 2 | 0 | 1 | 12 | 5 | |
| Ceuta | 2 | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 14 | 2 | 0 | |
| Ciudad Real | 3 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 2 | 4 | 1 | 4 | 4 | 1 | 1 | 4 | 12 | |
| Córdoba | 3 | 4 | 2 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 0 | 3 | 13 | 2 | |
| Coruña, A | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 16 | 0 | 0 | |
| Cuenca | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 6 | 0 | 0 | 12 | |
| Gipuzkoa | 2 | 3 | 3 | 3 | 4 | 2 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 0 | 2 | 8 | 8 | |
| Girona | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 0 | 0 | 17 | |
| Granada | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 8 | 10 | 0 | |
| Guadalajara | 4 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 1 | 4 | 1 | 1 | 1 | 4 | 1 | 4 | 10 | 0 | 0 | 8 | |
| Huelva | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 0 | 6 | 12 | 0 | |
| Huesca | 4 | 1 | 1 | 4 | 1 | 4 | 4 | 1 | 1 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 4 | 1 | 9 | 0 | 0 | 9 | |
| Jaén | 4 | 4 | 3 | 3 | 4 | 1 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 1 | 6 | 8 | 3 | |
| León | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 16 | 1 | 0 | 1 | |
| Lleida | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 4 | 5 | 0 | 13 | |
| Lugo | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 17 | 1 | 0 | 0 | |
| Madrid | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 0 | 0 | 12 | |
| Málaga | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 0 | 13 | 5 | 0 | |
| Melilla | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 1 | 2 | 1 | 2 | 4 | 12 | 2 | 0 | |
| Murcia | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 0 | 17 | 1 | |
| Navarra | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 0 | 0 | 17 | |
| Ourense | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 17 | 1 | 0 | 0 | |
| Palencia | 1 | 4 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 12 | 5 | 0 | 1 | |
| Palmas, Las | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 18 | 0 | 0 | |
| Pontevedra | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 16 | 1 | 0 | |
| Rioja, La | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 1 | 4 | 4 | 4 | 1 | 4 | 4 | 0 | 0 | 14 | |
| Salamanca | 4 | 4 | 1 | 4 | 1 | 4 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 12 | 0 | 0 | 6 | |
| Santa Cruz de Tenerife | 3 | 2 | 3 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 13 | 2 | 0 | |
| Segovia | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 15 | 0 | 0 | 3 | |
| Sevilla | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 0 | 4 | 14 | 0 | |
| Soria | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 17 | 0 | 0 | 1 | |
| Tarragona | 4 | 1 | 4 | 4 | 4 | 4 | 1 | 4 | 1 | 4 | 1 | 4 | 4 | 4 | 1 | 1 | 4 | 6 | 0 | 0 | 0 | 12 | |
| Teruel | 4 | 4 | 4 | 4 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 0 | 0 | 6 | |
| Toledo | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 18 | |
| Valencia/València | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 0 | 2 | 16 | 0 | |
| Valladolid | 4 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 5 | 0 | 0 | 0 | 13 | |
| Zamora | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 16 | 0 | 0 | 2 | |
| Zaragoza | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 0 | 0 | 2 | 16 | |

Fuente: Elaboración propia.

Según los resultados de la Tabla 3.7, se puede representar gráficamente la situación de las diferentes provincias españolas en relación con su esperanza de vida en general para los varones, teniendo en cuenta los grupos que definen los cuatro cuadrantes de los análisis de componentes principales realizados en cada año. Esta representación gráfica se muestra en la Figura 3.22.

Figura 3.22. Mortalidad provincial por tipo de mortalidad de acuerdo con los cuadrantes del plano de las dos primeras componentes principales. Varones.



Fuente: Elaboración propia.

Si realizamos el mismo tipo de análisis para los datos de la esperanza de vida por grupos quinquenales para las mujeres, durante los años que van desde 1991 hasta 2008, se obtienen resultados similares, por lo que se omite en este punto la representación gráfica de las dos primeras componentes principales en cada uno de los años considerados.

En la Tabla 3.8, se muestra la clasificación en cuadrantes de cada una de las provincias en cada año, atendiendo a la esperanza de vida de las mujeres. Las 4 columnas de la derecha proporcionan la frecuencia absoluta de veces que cada provincia aparece en cada cuadrante.

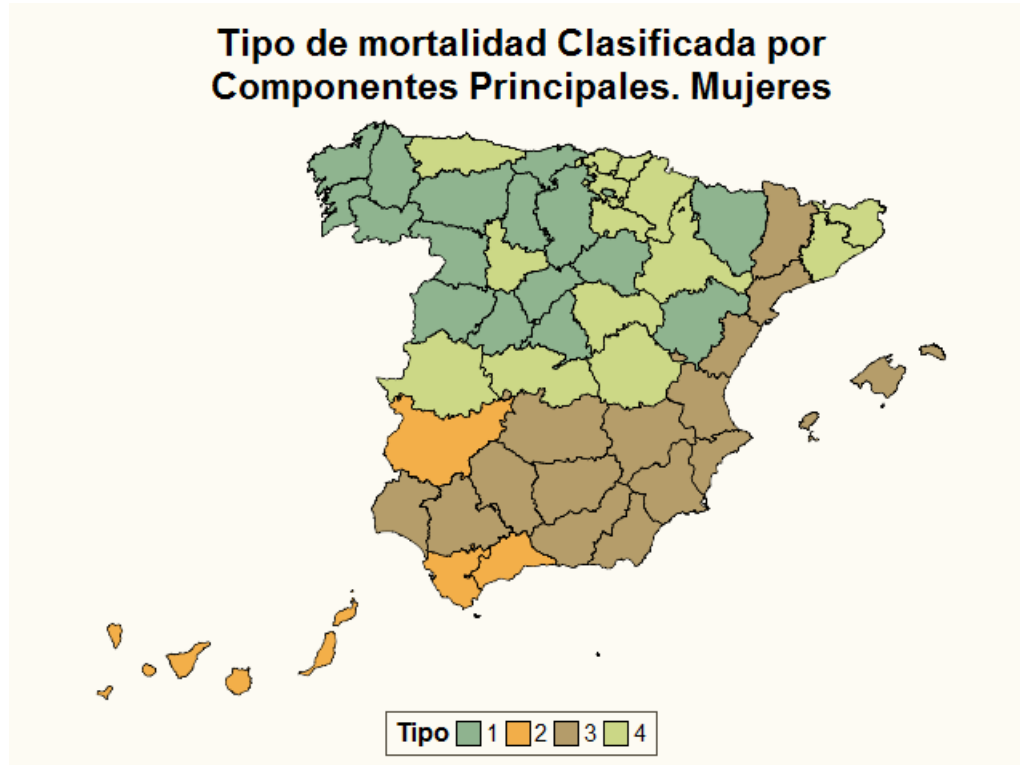
Tabla 3.8. Clasificación de las provincias por cuadrante por año y frecuencias absolutas. Mujeres.

| | año | | | | | | | | | | | | | | | | | | Número de | | | |
|------------------------|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----------|----|----|----|
| | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 1 | 2 | 3 | 4 |
| | Albacete | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 4 | 0 | 0 | 12 |
| Alicante/Alacant | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 0 | 4 | 14 | 0 |
| Almería | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 3 | 0 | 6 | 12 | 0 |
| Araba/Álava | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 9 | 0 | 0 | 9 | 0 |
| Asturias | 4 | 3 | 2 | 1 | 3 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 2 | 3 | 3 | 1 | 2 | 7 | 8 |
| Ávila | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 17 | 0 | 0 | 1 | 1 |
| Badajoz | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 0 | 11 | 7 | 0 |
| Balears, Illes | 3 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 1 | 4 | 3 | 1 | 4 | 2 | 3 | 11 | 2 |
| Barcelona | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 8 | 0 | 0 | 10 | 0 |
| Bizkaia | 4 | 1 | 4 | 1 | 4 | 4 | 4 | 1 | 4 | 1 | 4 | 1 | 1 | 4 | 4 | 1 | 1 | 9 | 0 | 0 | 9 | 0 |
| Burgos | 4 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 13 | 0 | 0 | 5 | 8 |
| Cáceres | 2 | 4 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 1 | 1 | 1 | 8 | 8 |
| Cádiz | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 0 | 15 | 3 | 0 |
| Cantabria | 1 | 4 | 4 | 1 | 1 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 12 | 0 | 0 | 6 | 6 |
| Castellón/Castelló | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 0 | 18 | 0 | 0 |
| Ceuta | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 18 | 0 | 0 | 0 |
| Ciudad Real | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 0 | 18 | 0 | 0 |
| Córdoba | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 1 | 17 | 0 | 0 |
| Coruña, A | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 14 | 3 | 0 | 1 | 1 |
| Cuenca | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 1 | 4 | 3 | 0 | 0 | 15 |
| Gipuzkoa | 1 | 4 | 4 | 1 | 4 | 1 | 4 | 4 | 4 | 1 | 4 | 1 | 4 | 4 | 4 | 4 | 1 | 6 | 0 | 0 | 12 | 0 |
| Girona | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 1 | 0 | 1 | 16 | 0 |
| Granada | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 0 | 2 | 16 | 0 |
| Guadalajara | 4 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 1 | 4 | 1 | 4 | 4 | 1 | 4 | 1 | 4 | 8 | 0 | 0 | 10 | 0 |
| Huelva | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 2 | 2 | 2 | 3 | 2 | 3 | 0 | 6 | 12 | 0 | 0 |
| Huesca | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 4 | 1 | 4 | 13 | 0 | 0 | 5 | 5 |
| Jaén | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 0 | 18 | 0 | 0 |
| León | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 14 | 0 | 0 | 4 | 3 |
| Lleida | 3 | 3 | 1 | 2 | 2 | 3 | 2 | 1 | 1 | 1 | 4 | 3 | 4 | 1 | 3 | 1 | 4 | 6 | 3 | 6 | 3 | 3 |
| Lugo | 2 | 2 | 4 | 1 | 4 | 1 | 4 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 11 | 2 | 0 | 5 | 5 |
| Madrid | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 4 | 4 | 4 | 1 | 14 | 0 | 0 | 4 | 0 |
| Málaga | 2 | 2 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 0 | 12 | 6 | 0 | 4 |
| Melilla | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 4 | 14 | 0 | 0 | 0 |
| Murcia | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 0 | 18 | 0 | 0 |
| Navarra | 1 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 4 | 7 | 0 | 0 | 11 | 0 |
| Ourense | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 15 | 0 | 0 | 3 | 3 |
| Palencia | 1 | 1 | 1 | 4 | 4 | 4 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 13 | 0 | 0 | 5 | 5 |
| Palmas, Las | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 18 | 0 | 0 | 0 |
| Pontevedra | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 13 | 4 | 0 | 1 | 1 |
| Rioja, La | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 18 | 0 |
| Salamanca | 1 | 1 | 4 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 16 | 0 | 0 | 2 | 2 |
| Santa Cruz de Tenerife | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 16 | 1 | 0 | 0 |
| Segovia | 1 | 4 | 4 | 1 | 4 | 1 | 4 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 | 0 | 0 | 5 | 5 |
| Sevilla | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 3 | 15 | 0 | 0 |
| Soria | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 16 | 0 | 0 | 2 | 2 |
| Tarragona | 3 | 1 | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 3 | 2 | 11 | 3 | 3 |
| Teruel | 4 | 1 | 4 | 4 | 1 | 4 | 4 | 1 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 0 | 0 | 8 | 8 |
| Toledo | 4 | 3 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 3 | 0 | 0 | 7 | 11 | 11 |
| Valencia/València | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 0 | 2 | 16 | 0 | 0 |
| Valladolid | 4 | 1 | 1 | 4 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 4 | 4 | 1 | 4 | 7 | 0 | 0 | 11 | 11 |
| Zamora | 1 | 1 | 4 | 1 | 4 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 14 | 0 | 0 | 4 | 4 |
| Zaragoza | 4 | 4 | 4 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 1 | 2 | 0 | 1 | 15 | 15 |

Fuente: Elaboración propia.

Según los resultados de la Tabla 3.8, se puede representar gráficamente la situación de las diferentes provincias españolas en relación con su esperanza de vida para las mujeres en general, teniendo en cuenta los grupos que definen los cuatro cuadrantes de los análisis de componentes principales realizados en cada año. Esta representación gráfica se muestra en la Figura 3.23.

Figura 3.23. Mortalidad provincial por tipo de mortalidad de acuerdo con los cuadrantes del plano de las dos primeras componentes principales. Mujeres.



Fuente: Elaboración propia.

De acuerdo con la Figura 3.23 se distinguen distintas regiones. La zona sur y mediterránea de tipo 2 y 3 y la zona norte y oeste de tipo 1 y 4 coincidiendo con análisis anteriores sobre la esperanza de vida.

El análisis por año proporciona resultados interesantes. Ahora, más deseable aún es disponer de la información en un mismo plano y así poder observar la evolución de las provincias a lo largo del tiempo. Este es el objeto del apartado que se presenta a continuación.

3.4.2. Análisis combinado

Para poder realizar un estudio de carácter longitudinal, y teniendo en cuenta la alta variabilidad que recogen las dos componentes extraídas en cada año analizado (ver Tabla 3.6), la primera posibilidad es suponer que la estructura de correlación entre

las variables se mantiene a lo largo del tiempo. En este caso, se podría aplicar un ACP sobre las 19 variables, considerando conjuntamente todos los años del período 1991-2008, utilizando la matriz de correlaciones combinada. De esta forma, se obtendrían dos componentes principales globales, que resumirían la información de todas las variables durante todo el período, con lo que se podrían establecer comparaciones transversales y longitudinales, al ser la base de referencia común. Por tanto, es necesario comprobar si la estructura de correlaciones que liga dichos indicadores es la misma en cada periodo. Para realizar dicha comprobación, se utiliza el contraste de la M de Box (para los detalles teóricos, véase Rencher, 1995, pág. 280 y siguientes), que permite comprobar la igualdad de matrices de varianzas y covarianzas, sobre las variables tipificadas.

Este contraste se formula teniendo en cuenta el determinante de las matrices de correlación, cuya información se muestra en la Tabla 3.9:

Tabla 3.9 Logaritmo del determinante de la matriz de correlaciones para las variables que miden la esperanza de vida por edades. Varones y mujeres

| Año | Logaritmo del Determinante Caso de Varones | Logaritmo del Determinante Caso de Mujeres |
|-------------------------------|--|--|
| 1991 | -76.866 | -80.620 |
| 1992 | -78.261 | -76.247 |
| 1993 | -77.799 | -80.489 |
| 1994 | -77.063 | -82.225 |
| 1995 | -75.829 | -79.108 |
| 1996 | -77.829 | -79.634 |
| 1997 | -76.480 | -83.460 |
| 1998 | -79.154 | -84.107 |
| 1999 | -77.996 | -85.512 |
| 2000 | -79.619 | -86.211 |
| 2001 | -79.820 | -85.539 |
| 2002 | -77.985 | -85.898 |
| 2003 | -80.039 | -86.046 |
| 2004 | -77.855 | -87.521 |
| 2005 | -77.235 | -86.929 |
| 2006 | -75.058 | -83.349 |
| 2007 | -76.429 | -85.198 |
| 2008 | -78.086 | -82.428 |
| Intra-grupos combinada | -72.889 | -77.379 |

Fuente: Elaboración propia a partir de datos del INE

En ninguno de los dos casos (varones o mujeres, véanse las Tablas 3.10 y 3.11) se puede asumir como cierta la hipótesis nula de igualdad de matrices de correlación

entre ambos grupos de variables a lo largo de los 18 años considerados, razón por la que no es posible encontrar un espacio bidimensional derivado mediante el ACP sobre la matriz de correlaciones combinada.

Tabla 3.10. y Tabla 3.11. Resultado de la Prueba de Box para las matrices de correlaciones de los indicadores de esperanza de vida. Varones tabla de arriba y mujeres de abajo.

| | | |
|----------|--------|------------|
| M de Box | | 4545,072 |
| F | Aprox. | 1,737 |
| | gl1 | 2312 |
| | gl2 | 441436,612 |
| | Sig. | ,000 |

| | | |
|----------|--------|------------|
| M de Box | | 1776,636 |
| F | Aprox. | 1,762 |
| | gl1 | 935 |
| | gl2 | 449072,500 |
| | Sig. | ,000 |

Fuente: Elaboración propia a partir de datos del INE

Una posible solución para poder realizar el análisis longitudinal de la información resumida de las variables consideradas, es la aplicación de la técnica del Análisis del Espacio Común (Krzanowski, 1979). Esta técnica es un método algebraico que permite encontrar un espacio bidimensional lo más próximo posible a todos los subespacios bidimensionales generados por el ACP realizado año por año. Este espacio común permitirá la comparación longitudinal y transversal de las provincias españolas en términos del comportamiento y la evolución de su esperanza de vida a las distintas edades.

La aplicación del espacio común parte de los dos vectores propios que definen las componentes principales en el ACP de cada año. Si se denomina U_i a la matriz de dimensión 18×2 cuyas columnas son los dos vectores propios de la matriz de correlación entre las variables consideradas en el año i desde 1991 hasta 2008, entonces la descomposición espectral de la matriz

$$H = \sum_{i=1991}^{2008} U_i \cdot U_i'$$

genera un subespacio de dimensión 2 que es el más próximo posible a los 18 espacios bidimensionales derivados del ACP, en el sentido de que cada una de las componentes

del espacio común minimiza la distancia a cada una de las componentes de los espacios bidimensionales generados para cada año.

En este caso, el método del análisis del espacio común es muy adecuado, puesto que no presenta los problemas que se le han atribuido habitualmente. En primer lugar, los datos analizados son muy estables en el tiempo. En cada instante temporal analizado, las dos componentes principales son muy estables, puesto que los dos primeros autovalores de las matrices de correlación son muy grandes en relación al resto (las dos componentes acumulan más de un 97% de la variabilidad observada en cualquier caso). En segundo lugar, estamos analizando en cada uno de los períodos de tiempo considerados un mismo número de variables medidas sobre el mismo conjunto de casos (las 52 provincias consideradas y el Total Nacional), con lo que no se encuentran irregularidades en los tamaños muestrales analizados.

Se aplica el análisis del espacio común, y se obtienen los siguientes coeficientes para el cálculo de las componentes del espacio común (por variables ordenadas) para varones, Tabla 3.12 y mujeres, Tabla 3.13:

Tabla 3.12. Coeficientes para las componentes del espacio común. Varones

| Variable | Peso en la Primera Componente | Peso en la Segunda Componente |
|----------|-------------------------------|-------------------------------|
| Zev00 | 0.22932 | -0.04126 |
| Zev01 | 0.23148 | -0.03717 |
| Zev05 | 0.23228 | -0.03700 |
| Zev10 | 0.23219 | -0.03675 |
| Zev15 | 0.23250 | -0.03625 |
| Zev20 | 0.23462 | -0.03547 |
| Zev25 | 0.23433 | -0.03190 |
| Zev30 | 0.23590 | -0.02971 |
| Zev35 | 0.23306 | -0.02659 |
| Zev40 | 0.22968 | -0.02265 |
| Zev45 | 0.22892 | -0.01771 |
| Zev50 | 0.22683 | -0.01052 |
| Zev55 | 0.22552 | -0.00132 |
| Zev60 | 0.22269 | 0.00748 |
| Zev65 | 0.21641 | 0.02355 |
| Zev70 | 0.21519 | 0.04674 |
| Zev75 | 0.21670 | 0.07677 |
| Zev80 | 0.19900 | 0.12867 |
| Zev85 | 0.15028 | 0.16973 |

Fuente: Elaboración propia a partir de datos del INE

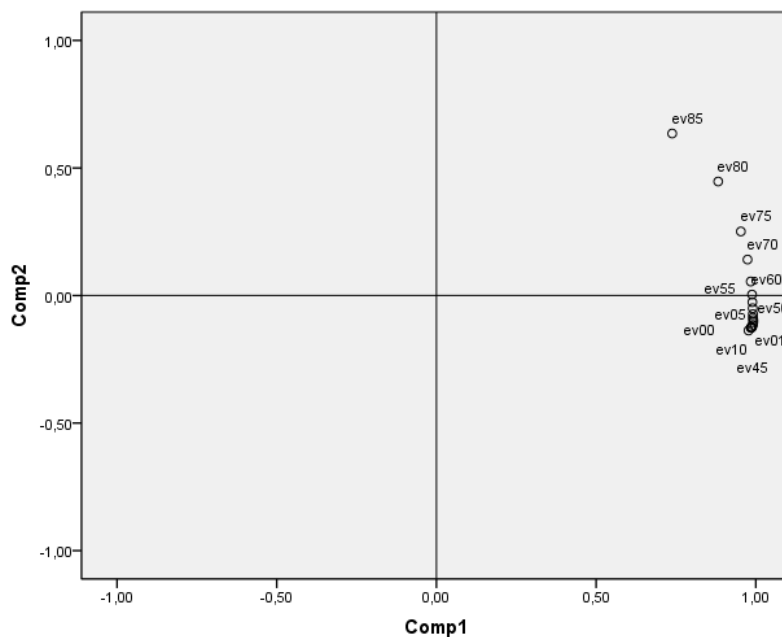
Tabla 3.13: Coeficientes para las componentes del espacio común. Mujeres

| Variable | Peso en la Primera Componente | Peso en la Segunda Componente |
|----------|-------------------------------|-------------------------------|
| Zem00 | 0,23596 | -0,03616 |
| Zem01 | 0,23374 | -0,03053 |
| Zem05 | 0,23483 | -0,02964 |
| Zem10 | 0,23282 | -0,02857 |
| Zem15 | 0,23359 | -0,02762 |
| Zem20 | 0,23492 | -0,02706 |
| Zem25 | 0,23560 | -0,02616 |
| Zem30 | 0,23251 | -0,02395 |
| Zem35 | 0,23276 | -0,02203 |
| Zem40 | 0,23288 | -0,01942 |
| Zem45 | 0,23190 | -0,01634 |
| Zem50 | 0,23335 | -0,01015 |
| Zem55 | 0,23122 | -0,00742 |
| Zem60 | 0,22854 | 0,00195 |
| Zem65 | 0,22571 | 0,01508 |
| Zem70 | 0,21863 | 0,03270 |
| Zem75 | 0,21384 | 0,05903 |
| Zem80 | 0,20907 | 0,09588 |
| Zem85 | 0,19478 | 0,14217 |

Fuente: Elaboración propia a partir de datos del INE

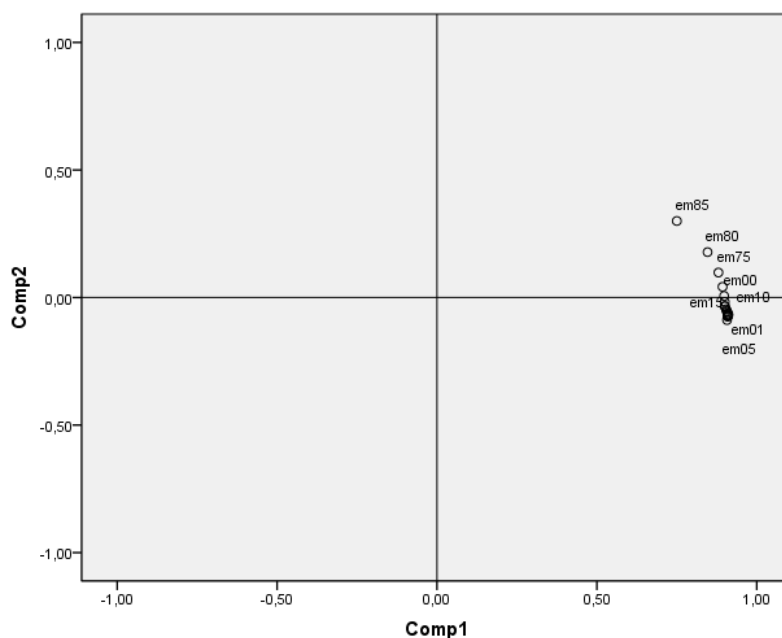
Una vez aplicado el método del Análisis del Espacio Común, se obtiene una nueva configuración bidimensional de la información de la esperanza de vida a las distintas edades consideradas para varones y mujeres durante los 18 años considerados. Para entender exactamente qué explica cada una de estas dos componentes comunes, se presentan en las Figuras 3.24 y 3.25 los coeficientes de correlación entre las variables observadas y las dos Componentes Comunes obtenidas, a modo de Gráfico de Componentes en un Análisis de Componentes Principales, diferenciando los casos de varones y mujeres que, como se ha indicado en el apartado introductorio, presentan diferencias entre sí. Hay que observar que en este análisis, ya se están considerando conjuntamente todas las provincias a lo largo de todo el período temporal analizado, lo que permitirá representar todos los casos de análisis, facilitando la comparación tanto transversal como longitudinal de los mismos.

Figura 3.24: Coeficientes de correlación entre las variables observadas y las Componentes Comunes. Varones.



Fuente: Elaboración propia a partir de datos del INE

Figura 3.25: Coeficientes de correlación entre las variables observadas y las Componentes Comunes. Mujeres.



Fuente: Elaboración propia a partir de datos del INE

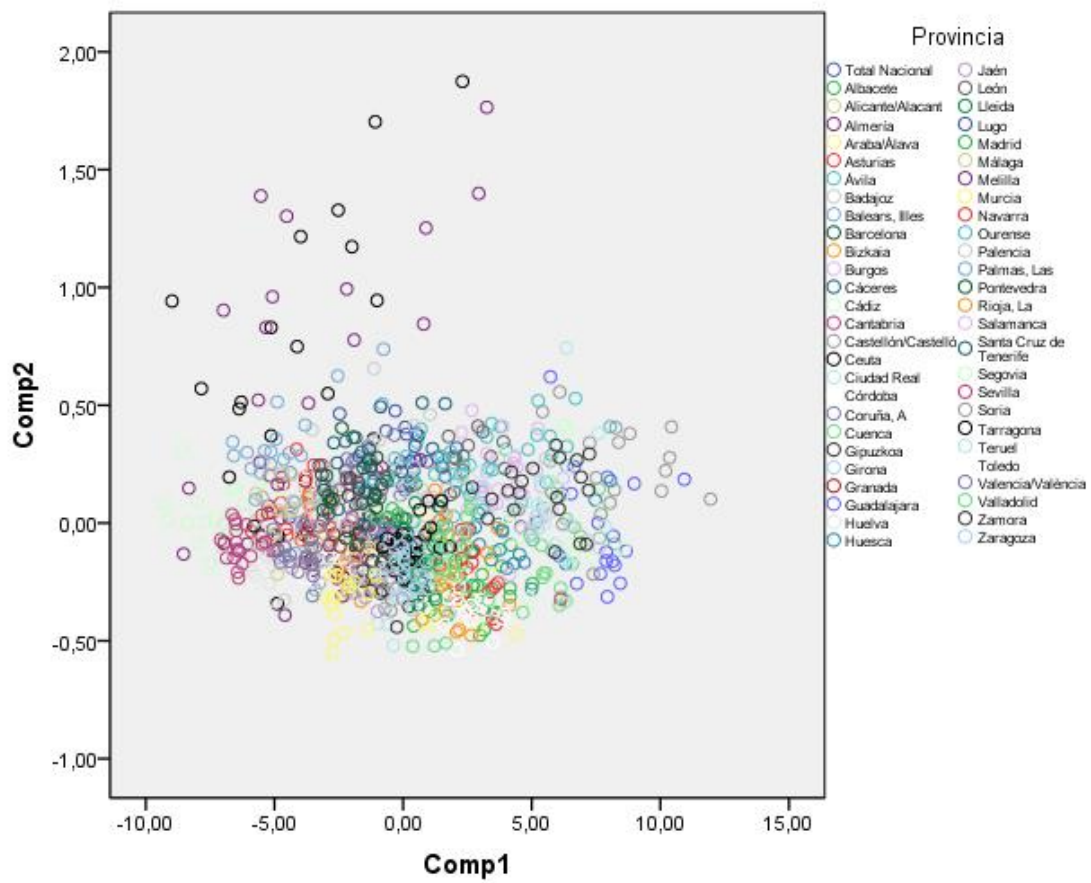
Como se puede apreciar a partir de las Figuras 3.24 y 3.25, la primera componente del Espacio Común es una componente de tamaño, pues se encuentra

altamente correlacionada con todas las variables analizadas. De este modo, aquellas provincias que alcancen un valor alto en esta componente serán las que tengan un mejor comportamiento en cuanto a la esperanza de vida que presentan a las distintas edades, mientras que las que obtengan un valor más bajo, tendrán una peor perspectiva en cuanto a los años que restan por vivir a las diferentes edades. De este modo, se puede denominar a esta primera componente del espacio común como *expectativa de vida*.

La segunda componente se relaciona escasamente con las esperanzas de vida a las diversas edades, aunque se aprecia una ligera correlación positiva con la esperanza de vida a edades más altas, como es el caso de los varones, en el que la correlación entre la segunda componente del espacio común y la esperanza de vida a los 85 años es únicamente de 0,615. Por esta razón, no parece lógico emplearla en el estudio, puesto que la información que aporta no es demasiado relevante.

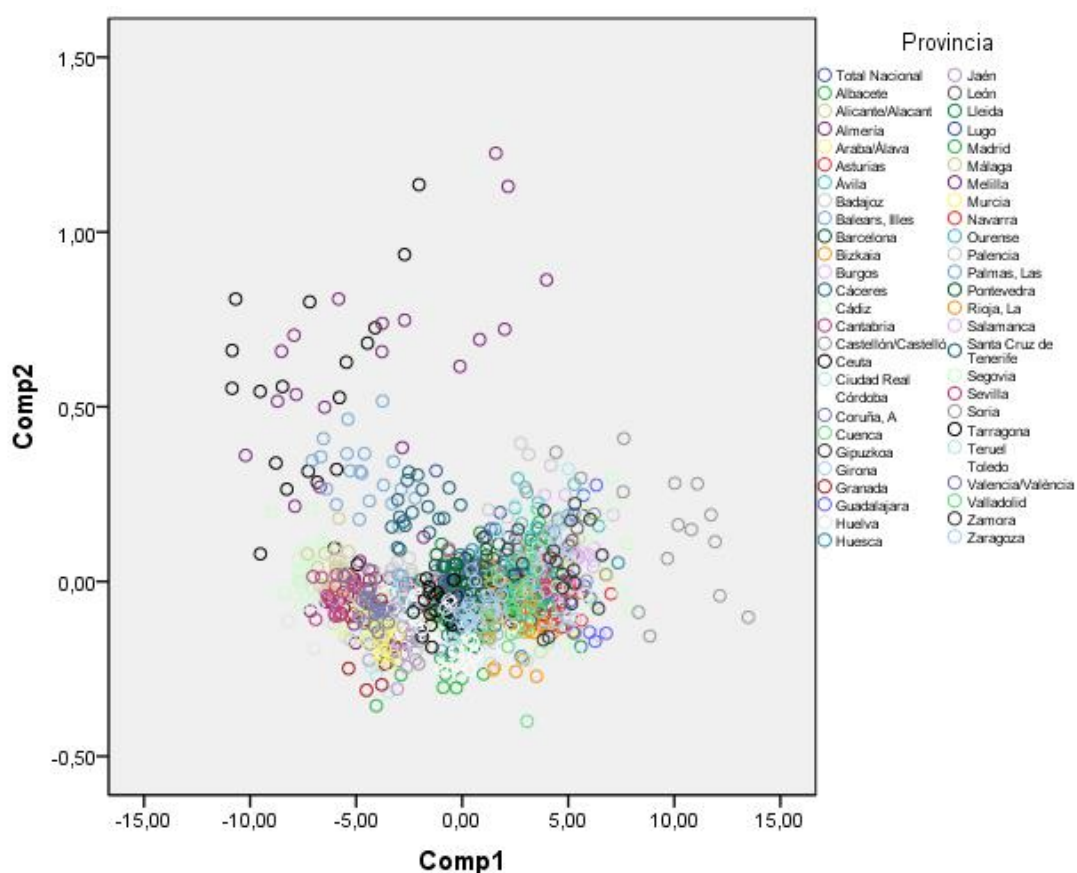
En las Figuras 3.26 y 3.27 se representan las puntuaciones de todas las provincias y las Ciudades Autónomas en el espacio común obtenido. Por tanto se incluyen para cada una de ellas los 18 puntos que sitúan su evolución a lo largo de todo el período 1991-2008. Así, los casos que quedan más a la derecha tienen una expectativa de vida superior al Total Nacional, que se encuentra aproximadamente en la parte central de la gráfica, mientras que los casos que quedan más a la izquierda poseen una expectativa de inferior al Total Nacional.

Figura 3.26. Representación de las provincias en las componentes del Espacio Común. Varones.



Fuente: Elaboración propia a partir de datos del INE

Figura 3.27. Representación de las provincias en las componentes del Espacio Común. Mujeres.



Fuente: Elaboración propia a partir de datos del INE

A continuación se hace un análisis pormenorizado de la trayectoria seguida por cada provincia con relación a la seguida por el Total Nacional, con el objetivo de caracterizar a las provincias en tres grupos. El primer grupo estará formado por aquellas que tengan una menor esperanza de vida que el Total Nacional, y presentaran valores más pequeños que el Total Nacional en la Primera Componente Común durante todo el período de análisis. El segundo grupo, estará formado por aquellas provincias que tengan en general una mayor esperanza de vida a las distintas edades que el Total Nacional en todo el período, y sus trayectorias en el Espacio Común se situarán a la derecha. Por último, incluiremos en un grupo intermedio a las provincias que no permanezcan siempre por encima o por debajo del Total Nacional, sino que sus trayectorias se crucen con él a lo largo del tiempo.

En la Tabla 3.14, se muestra la clasificación obtenida según el criterio considerado para comparar la trayectoria seguida por cada provincia en relación con el Total Nacional, para el caso de los varones.

Tabla 3.14. Clasificación de las provincias españolas atendiendo a su comportamiento general en cuanto a la expectativa de vida en relación con el Total Nacional. Varones.

| Provincias con peor expectativa de vida que el Total Nacional | Provincias con expectativa de vida no comparable con el Total Nacional | Provincias con mejor expectativa de vida que el Total Nacional |
|---|--|---|
| Almería Asturias Badajoz Cádiz Granada Huelva Las Palmas Málaga Sevilla Valencia | Alicante Araba/Álava Balears, Illes Barcelona Bizkaia Cáceres Cantabria Castellón/Castelló Ceuta Ciudad Real Córdoba Coruña, A Guipuzkoa Jaén León Lugo Melilla Murcia Ourense Palencia Pontevedra Santa Cruz de Tenerife Zaragoza | Albacete Ávila Burgos Cuenca Girona Guadalajara Huesca La Rioja Lleida Madrid Navarra Salamanca Segovia Soria Tarragona Teruel Toledo Valladolid Zamora |

Fuente: Elaboración propia a partir de datos del INE

En la Tabla 3.15, se muestra la situación de cada una de las provincias españolas en relación a la evolución del Total Nacional en cuanto a la expectativa de vida obtenida según el método del Espacio Común para las mujeres

Tabla 3.15. Clasificación de las provincias españolas atendiendo a su comportamiento general en cuanto a la expectativa de vida en relación con el Total Nacional. Mujeres.

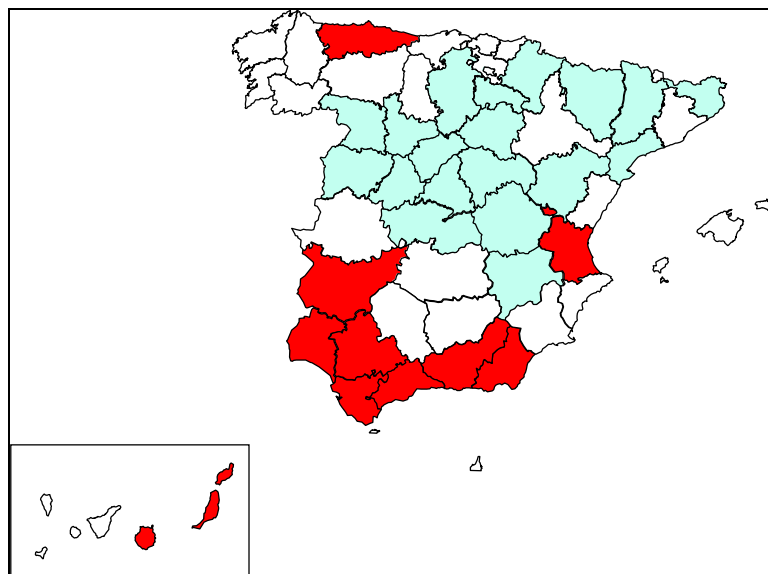
| Provincias con peor expectativa de vida que el Total Nacional | Provincias con expectativa de vida no comparable con el Total Nacional | Provincias con mejor expectativa de vida que el Total Nacional |
|--|--|--|
| Alicante Almería Badajoz Cádiz Castellón/Castelló Ceuta Ciudad Real Córdoba Granada Huelva Jaén Las Palmas Málaga Murcia Sevilla Valencia | Albacete Asturias Balears, Illes Cáceres Coruña, A Girona Lleida Lugo Melilla Pontevedra Santa Cruz de Tenerife Tarragona Toledo Zaragoza | Araba/Álava Ávila Barcelona Bizkaia Burgos Cantabria Cuenca Guadalajara Guipuzkoa Huesca La Rioja León Madrid Navarra Ourense Palencia Salamanca Segovia Soria Teruel Valladolid Zamora |

Fuente: Elaboración propia a partir de datos del INE

En los grupos que se describen en las Tablas 3.14 y 3.15 se aprecia cierta estabilidad, y se encuentran algunas provincias cuya expectativa de vida es inferior a la del Total Nacional en cualquier caso (considerando los varones o las mujeres). Concretamente, se trata de las provincias de Almería, Badajoz, Cádiz, Granada, Huelva, Las Palmas, Málaga, Sevilla y Valencia, que forman un arco sur de territorios cuya expectativa de vida es inferior a la del Total Nacional. Por otra parte, hay provincias cuya expectativa de vida supera a la del Total Nacional en ambas subpoblaciones de hombres y mujeres. Se trata de las provincias de Zamora, Salamanca, Valladolid, Ávila, Segovia, Burgos, Soria, La Rioja, Navarra, Teruel, Huesca, Madrid, Guadalajara y Cuenca. Estas provincias forman un territorio en el que la expectativa de vida toma mayores valores que la del Total Nacional.

En la Figura 3.28 se muestran en azul las provincias cuya expectativa de vida es superior a la del Total Nacional en todos los años para el colectivo de varones, mientras que en rojo se señalan aquellas con valores más pequeños.

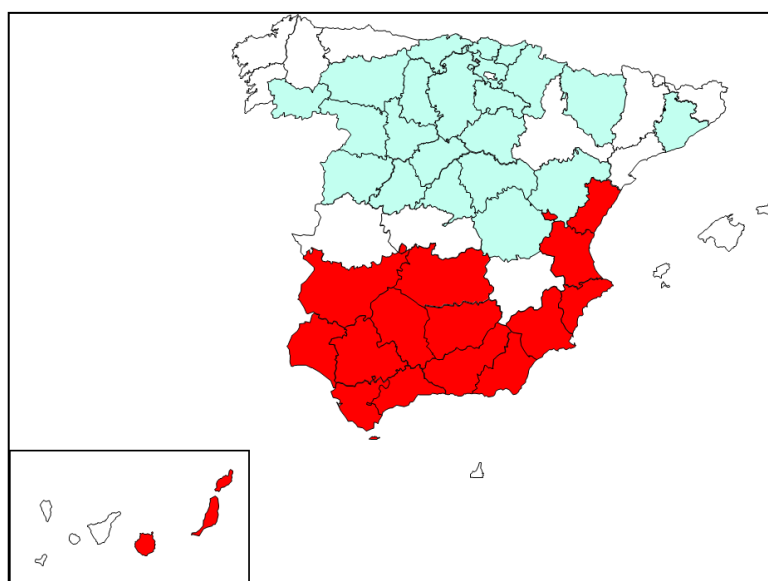
Figura 3.28. Grupos de provincias con mejor o peor expectativa de vida que el Total Nacional en el período 1991-2008. Varones.



Fuente: Elaboración propia a partir de datos del INE

En la Figura 3.29 se muestran sombreados en azul las provincias cuya expectativa de vida es superior a la del Total Nacional en todos los años para el colectivo de mujeres, mientras que en rojo se señalan aquellas con valores más pequeños.

Figura 3.29. Grupos de provincias con mejor o peor expectativa de vida que el Total Nacional en el período 1991-2008. Mujeres.

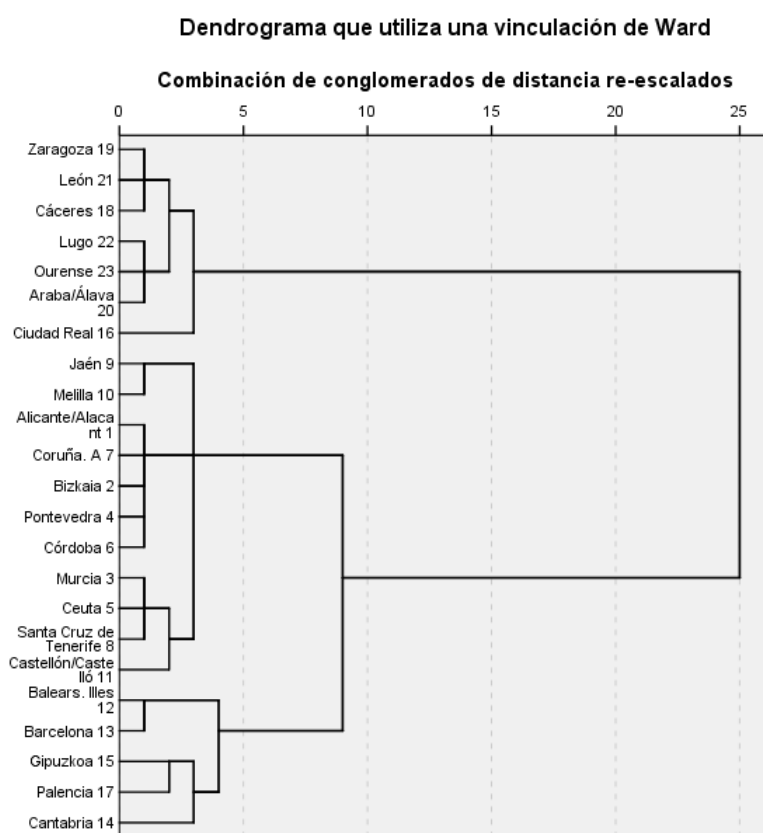


Fuente: Elaboración propia a partir de datos del INE

Otro aspecto que parece importante es analizar qué ocurre con las provincias que se sitúan en el grupo intermedio de provincias con expectativa no comparable con el Total Nacional, para tratar de refinar el conocimiento sobre la evolución de la trayectoria que ha seguido la esperanza de vida de cada provincia en relación con la del Total Nacional, que usamos como elemento de comparación. Para realizar este estudio, se ha aplicado un análisis de conglomerados usando el método de Ward, y tomando como medida de distancia entre los casos de análisis la distancia euclídea al cuadrado con datos binarios que indican si cada provincia está por encima o por debajo del Total Nacional en cada año considerado.

En el caso de los varones, el dendrograma obtenido por esta metodología para expresar esta clasificación se presenta en la Figura 3.30, en la que se puede apreciar con claridad la aparición de tres grupos bien definidos de provincias:

Figura 3.30: Dendrograma de clasificación de las provincias con Expectativa de Vida no comparable con el Total Nacional en el período 1991-2008. Varones.



Fuente: Elaboración propia.

Los grupos que aparecen en esta clasificación se muestran en la Tabla 3.16:

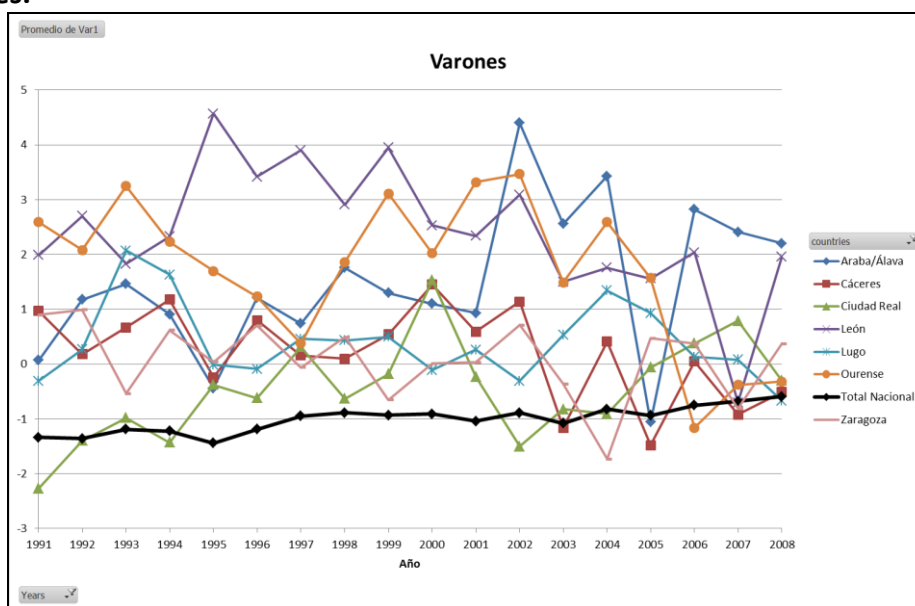
Tabla 3.16: Clasificación de las provincias con Expectativa de Vida no comparable con el Total Nacional en el período 1991-2008. Varones.

| Grupo A | Grupo B | Grupo C |
|-------------|------------------------|----------------|
| Zaragoza | Jaén | Balears, Illes |
| León | Melilla | Barcelona |
| Cáceres | Alicante/Alacant | Gipuzkoa |
| Lugo | Castellón/Castelló | Palencia |
| Ourense | Coruña, A | Cantabria |
| Araba/Álava | Bizkaia | |
| Ciudad Real | Pontevedra | |
| | Córdoba | |
| | Murcia | |
| | Ceuta | |
| | Santa Cruz de Tenerife | |

Fuente: Elaboración propia

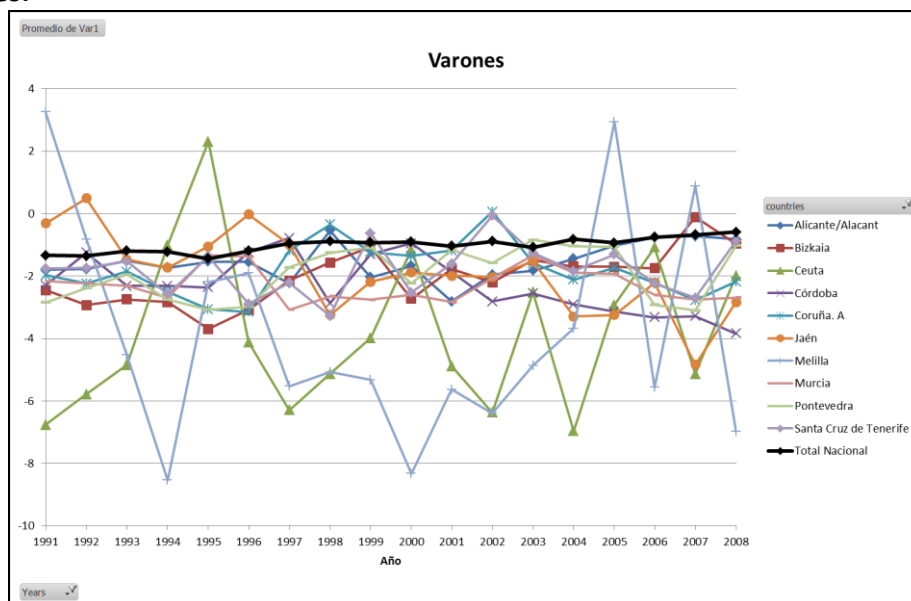
El Grupo A está formado por provincias que están casi siempre por encima del Total Nacional, el Grupo B contiene provincias cuyas trayectorias están casi siempre por debajo del Total Nacional, mientras que las provincias del Grupo C tienen trayectorias que se cruzan con la del Total Nacional a lo largo del período considerado. Para comprobarlo, en las Figuras 3.31, 3.32 y 3.33 se muestran las trayectorias de las provincias de cada grupo, junto con la del Total Nacional.

Figura 3.31: Trayectorias de las provincias del grupo A en el período 1991-2008. Varones.



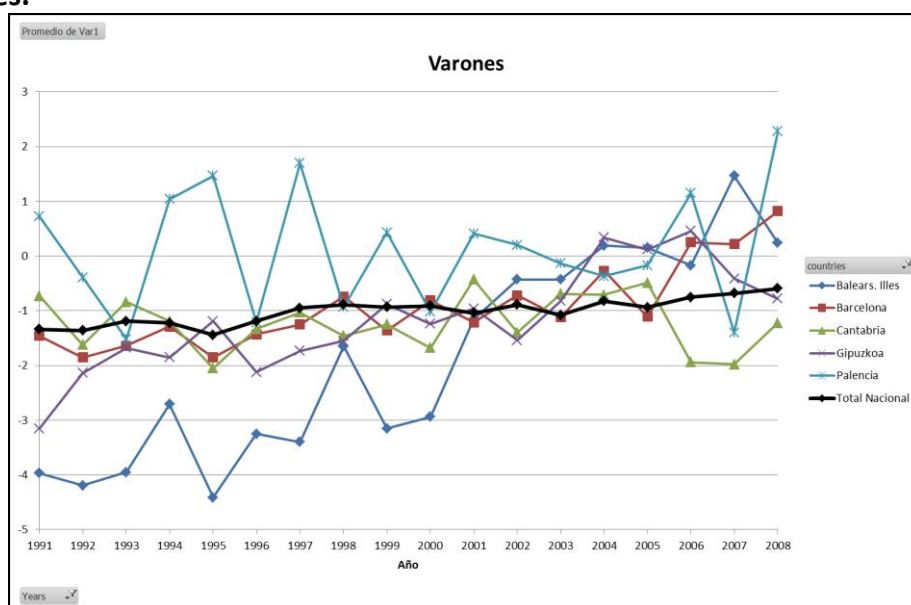
Fuente: Elaboración propia

Figura 3.32: Trayectorias de las provincias del grupo B en el período 1991-2008. Varones.



Fuente: Elaboración propia

Figura 3.33: Trayectorias de las provincias del grupo C en el período 1991-2008. Varones.

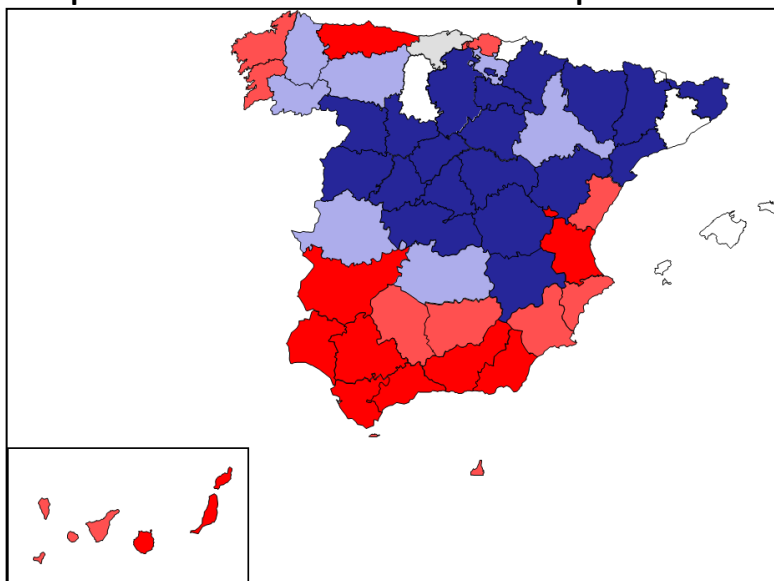


Fuente: Elaboración propia

De esta forma, podemos clasificar a las provincias en cinco grupos según sea el comportamiento de su expectativa de vida en relación con la del Total Nacional, en el caso de los varones, según se muestra en la Figura 3.34, en la que los casos en azul oscuro están siempre por encima del Total Nacional durante el período 1991-2008, en

azul celeste están casi siempre por encima del Total Nacional, los casos en rosa están casi siempre por debajo del Total Nacional, mientras que en rojo se muestran las provincias que están siempre por debajo del Total Nacional. Se dejan en blanco las provincias que presentan demasiadas oscilaciones con respecto del Total Nacional.

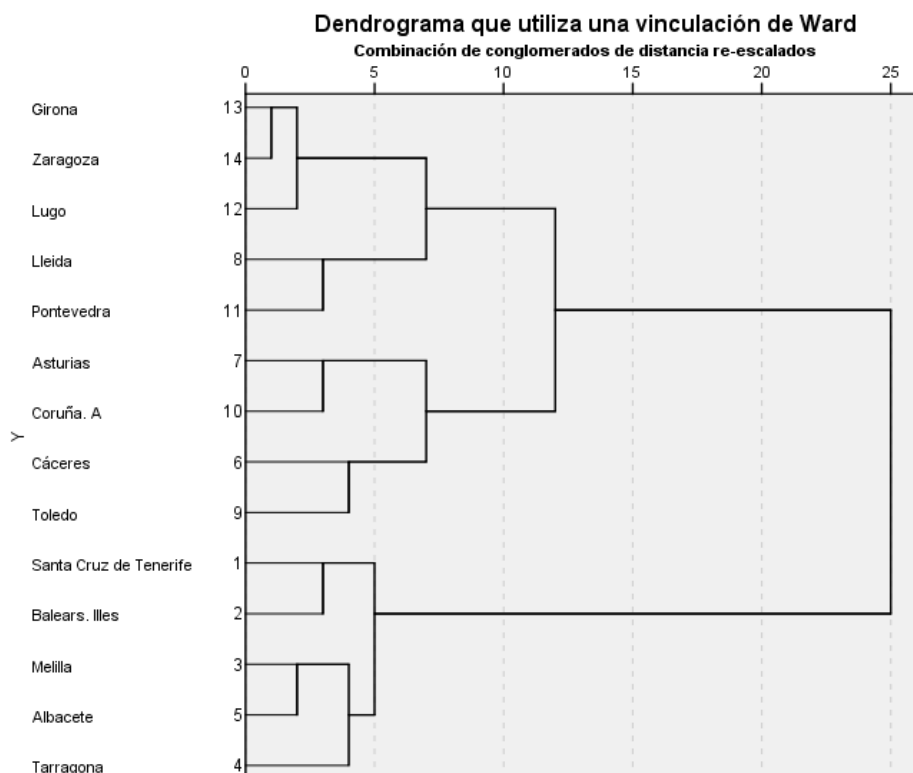
Figura 3.34. Clasificación 5 grupos de las provincias españolas según su expectativa de vida comparada con la del Total Nacional en el período 1991-2008. Varones.



Fuente: Elaboración propia

Aplicando esta misma metodología al análisis de la expectativa de vida de las mujeres en las provincias cuya expectativa de vida no es comparable con la del Total Nacional, y que se sitúan en el grupo intermedio de provincias que se muestra en la Tabla 3.15, se obtiene el dendrograma que se muestra en la Figura 3.35, en el que se vuelve a apreciar la aparición de tres grupos bien definidos.

Figura 3.35: Dendrograma de clasificación de las provincias con Expectativa de Vida no comparable con el Total Nacional en el período 1991-2008. Mujeres.



Fuente: Elaboración propia.

Los tres grupos que aparecen en esta clasificación se muestran en la Tabla 3.17:

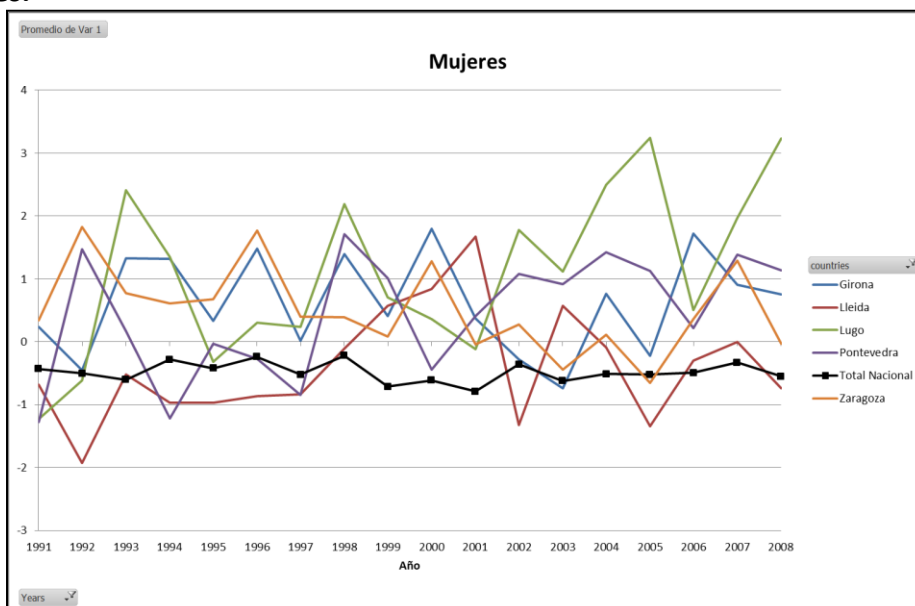
Tabla 3.17: Clasificación de las provincias con Expectativa de Vida no comparable con el Total Nacional en el período 1991-2008. Mujeres.

| Grupo A | Grupo B | Grupo C |
|--|--|--|
| Girona Zaragoza Lugo Lleida Pontevedra | Santa Cruz de Tenerife Balears, Illes Melilla Albacete Tarragona | Asturias Coruña, A Cáceres Toledo |

Fuente: Elaboración propia.

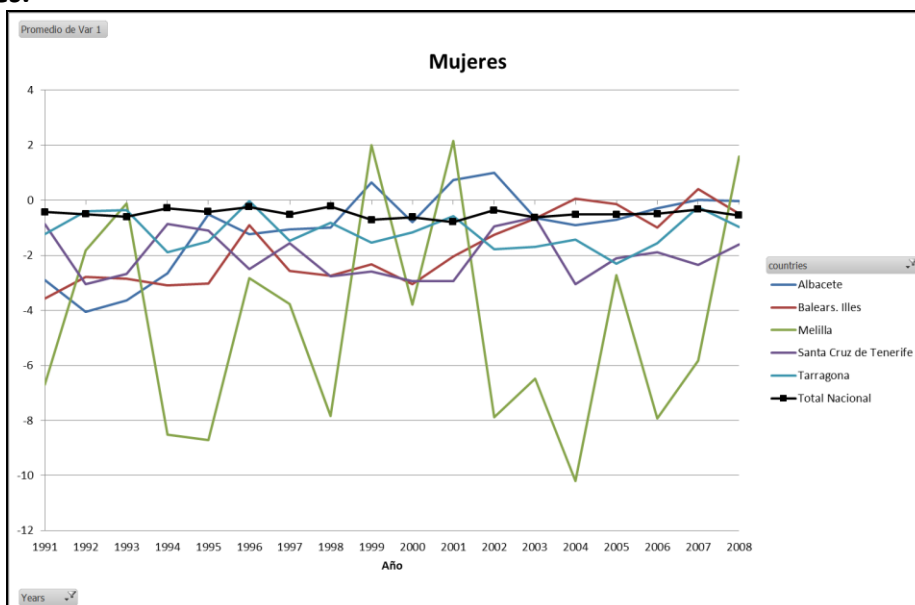
El Grupo A está formado por provincias que están casi siempre por encima del Total Nacional, el Grupo B contiene provincias cuyas trayectorias están casi siempre por debajo del Total Nacional, mientras que las provincias del Grupo C tienen trayectorias que se cruzan con la del Total Nacional a lo largo del período considerado. Para comprobarlo, en las Figuras 3.36, 3.37 y 3.38 se muestran las trayectorias de las provincias de cada grupo, junto con la del Total Nacional.

Figura 3.36: Trayectorias de las provincias del grupo A en el período 1991-2008. Mujeres.



Fuente: Elaboración propia.

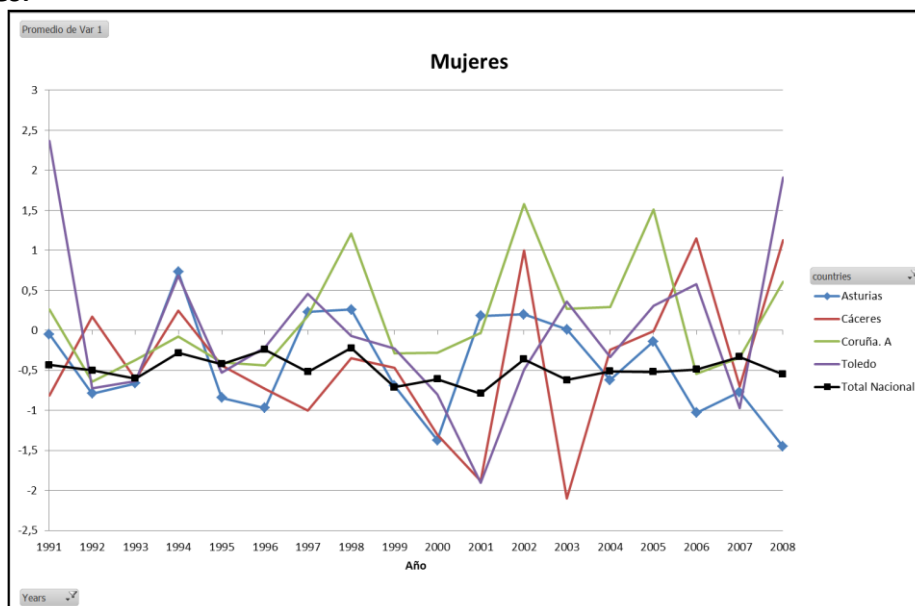
Figura 3.37: Trayectorias de las provincias del grupo B en el período 1991-2008. Mujeres.



Elaboración propia

Fuente:

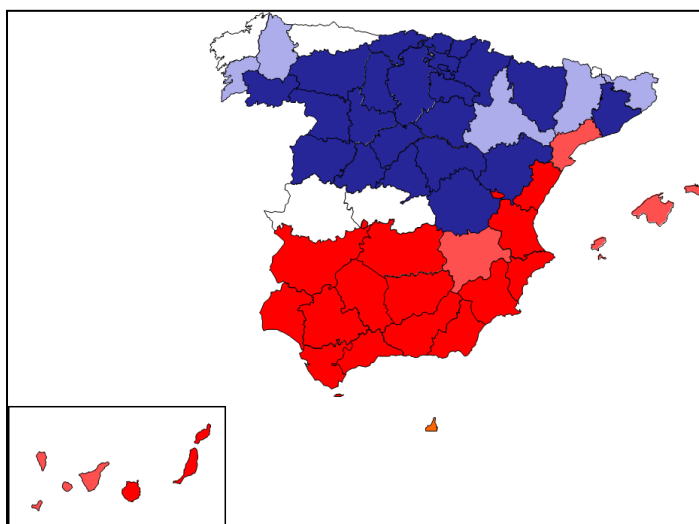
Figura 3.38: Trayectorias de las provincias del grupo C en el período 1991-2008. Mujeres.



Fuente: Elaboración propia

De esta forma, podemos clasificar a las provincias en cinco grupos según sea el comportamiento de su expectativa de vida en relación con la del Total Nacional, en el caso de las mujeres, según se muestra en la Figura 3.34, en la que los casos en azul oscuro están siempre por encima del Total Nacional durante el período 1991-2008, en azul celeste están casi siempre por encima del Total Nacional, los casos en rosa están casi siempre por debajo del Total Nacional, mientras que en rojo se muestran las provincias que están siempre por debajo del Total Nacional. Se dejan en blanco las provincias que presentan demasiadas oscilaciones con respecto del Total Nacional.

Figura 3.39. Clasificación 5 grupos de las provincias españolas según su expectativa de vida comparada con la del Total Nacional en el período 1991-2008. Mujeres.



Fuente: Elaboración propia.

De acuerdo con la Figura 3.34 y la Figura 3.39, que resumen todo el proceso de clasificación de acuerdo con el análisis combinado, se observa que la distribución geográfica relativa es similar para varones y para mujeres. La zona sur y las provincias costeras tienen peor mortalidad frente a la zona norte e interior siendo la discriminación más radical en las mujeres que en los varones. Se trata de una discrepancia en la incidencia de la mortalidad provincial que parece conservarse a lo largo del tiempo en contra de la hipótesis de convergencia territorial.

3.5 Conclusiones

En este capítulo se realiza un análisis descriptivo de las funciones biométricas de la tabla de mortalidad de España y sus provincias. De acuerdo con los resultados observados se confirma la estabilidad de los datos a nivel Nacional. En contraste, las provincias manifiestan una importante inestabilidad.

Este fenómeno es consecuencia directa de la naturaleza estimativa de los cálculos que son especialmente sensibles al tamaño de la muestra y a factores aleatorios coyunturales. Se pone en evidencia la naturaleza contable de las funciones biométricas, ya que se construyen a partir del registro individual en el MNP.

Todas las funciones biométricas de la tabla de mortalidad de España, tanto para varones como para mujeres, presentan una significativa contigüidad respecto a la edad y el tiempo. Además las funciones recursivas son más estables según aumente el tamaño de la población de partida a partir de la cual se estima el fenómeno.

En cuanto a cada una de las funciones para el total nacional, se observa que el riesgo de muerte y la tasa determinan superficies de mortalidad similares. Para la media del tiempo vivido el último año de vida se observa contigüidad en el tiempo en la edad, aparentemente menos clara por el detalle de la resolución y además se aprecian distintas regiones. El análisis de la superficie permite detectar patrones en ciertas generaciones. Estos patrones aparecen aparentemente sin motivo. Una posible explicación se detalla en el apartado 3.2 y viene asociada a la depuración sistemática de defunciones para las que se desconoce la generación y que se imputa a una generación terminada en cero. Esta hipótesis requiere confirmación y deja abierta una nueva línea de investigación. El resto de las superficies definidas por las funciones biométricas presentan una significativa contigüidad respecto a la edad y el tiempo, siendo más estables las funciones como la de supervivientes, la población estacionaria o la esperanza de vida que se construyen recursivamente frente a otras como las defunciones teóricas que presentan mayores oscilaciones.

Se confirma empíricamente que el máximo de estabilidad se alcanza para la esperanza de vida al nacimiento, indicador que resume la incidencia de la mortalidad como estimación a partir del total de la población. Otros indicadores que ganan en estabilidad son la función de supervivientes y la población estacionaria, en este caso según se sube en edad.

Frente a la continuidad de las superficies nacionales, las provincias presentan importantes fluctuaciones de año en año y de edad en edad. Esta característica se acentúa en las funciones no recursivas como el riesgo de muerte. La causa es la mayor sensibilidad a factores aleatorios y sucesos raros en estimaciones con poblaciones de pequeño tamaño. Los mapas de coropletas facilitan la comprensión de la distribución geográfica y permiten apreciar un comportamiento más anárquico donde es difícil descubrir patrones.

La observación de las series de las provincias a lo largo del tiempo permite apreciar un interesante resultado. El rango de la dispersión es estable en el tiempo tanto para el riesgo como para la esperanza de vida. Esta característica pone en duda la hipótesis de convergencia de la mortalidad hacia la tendencia de los territorios más desarrollados (Guijarro y Peláez, 2009), base de modelos clásicos de predicción. Por otra parte se aprecia que la amplitud del rango depende del tamaño de la variable del total nacional con una relación exponencial entre la pseudovarianza de las provincias y el tamaño de la variable. Para una detallada explicación al respecto, con una amplia descripción gráfica se puede consultar el artículo de López, Gutiérrez de Mesa y Rivera (2012).

Todas las funciones biométricas miden la incidencia de la mortalidad, pero cada una con sus matices. La esperanza de vida presenta mayor estabilidad que el riesgo de muerte, mayor según bajamos en edad y por tanto se incrementa el tamaño de la muestra a partir de la cual se estima el parámetro.

Tanto la esperanza de vida al nacimiento como la esperanza de vida a los 65 años tienen un comportamiento relativo entre provincias similar. Existe una significativa contigüidad entre provincias. Se aprecian distintas regiones. Una región con mortalidad más favorable que se extiende por el interior y el norte. Otra región con una mortalidad menos favorable que se extiende por las provincias costeras y el sur. Estas regiones son estables con ligeras evoluciones a lo largo del tiempo. La esperanza de vida es menor en las provincias de la costa frente a las interiores y cuanto más al sur y oeste frente al norte y este. Esta distribución sufre ligeros cambios a lo largo de los años en donde la zona de menor esperanza de vida sigue dominando el sur y se desplaza hacia el oeste.

Como extensión del resultado se abre una línea de investigación que conduce a una revisión conceptual de la esperanza de vida. Hasta la fecha, la bibliografía reconoce la esperanza de vida como una característica asociada a un territorio. Frente a este enfoque se introduce una innovadora concepción que vincula la esperanza de vida a un punto georreferenciado a partir del cual se construye la población que mejor explique la esperanza de vida del individuo teórico objeto de estudio. El fruto directo

de este razonamiento es la representación de la esperanza de vida como una superficie continua.

En el apartado 3.4.1 se procede a clasificar las provincias atendiendo a su mortalidad de acuerdo con un análisis de componentes principales de la serie de esperanzas de vida año a año. Después de una primera aproximación a través de la esperanza de vida al nacimiento, como indicador clásico y más representativo de la mortalidad, y la esperanza de vida a los 65 años, sigue un estudio exhaustivo de la mortalidad provincial teniendo en cuenta la esperanza de vida a cualquier edad. Se trata del correspondiente análisis de componentes principales. Para cada año la clasificación de las provincias da como resultado que la información se resume casi en su totalidad con la primera componente. Hecho por otra parte esperable por la construcción recursiva de la variable. Y añadiendo la segunda componentes se explica por encima del 98%. Es decir la mortalidad provincial se resume en dos indicadores como mucho.

Con la primera componente se establece una regla que clasifica las provincias en tres grupos principales según su mortalidad sea sistemáticamente más favorable que el total nacional, que incluye a las provincias del interior y el noreste en ambos sexos; menos favorable que el total nacional todos los años que incluye las provincias costeras y sureñas y hacia el oeste para los varones; y un tercer grupo intermedio en el que se clasifica el resto de provincias que cambia su comportamiento a lo largo del tiempo.

Aprovechando la segunda componente se sigue un análisis sobre el plano de las dos componentes principales. El total nacional determina cuatro cuadrantes en el plano de las componentes principales. A partir de la representación por rangos de las provincias se siguen observa que las provincias tienen un comportamiento significativamente estable y evolucionan en el tiempo de forma continua sin cambios bruscos de cuadrante tanto en varones como mujeres coincidiendo en gran parte los resultados. Este comportamiento, además, confirma que la mortalidad es significativamente independiente respecto al sexo en términos relativos por provincias.

Además se confirma que se distinguen dos regiones claramente diferenciadas, una región al sur y bordeando la costa, que presenta, sistemáticamente, peor mortalidad que el total nacional. Otra región al norte y más interior con mejor mortalidad que el total nacional.

Finalmente, para completar la clasificación de las provincias en el apartado 3.4.2 se analiza la mortalidad combinando todos los años aprovechando toda la información en el mismo plano y así observar la evolución de las provincias a lo largo del tiempo. Los resultados y conclusiones de nuevo inciden en la misma dirección que los demás análisis realizados. De acuerdo con el análisis combinado, se observa que la distribución geográfica relativa es similar para varones y para mujeres. La zona sur y las provincias costeras tienen peor mortalidad frente a la zona norte e interior siendo la discriminación más radical en las mujeres que en los varones. Se trata de una discrepancia en la incidencia de la mortalidad provincial que parece conservarse a lo largo del tiempo en contra de la hipótesis de convergencia territorial.

Del análisis combinado de todos los años con el método del Análisis del Espacio Común definido por dos componentes se concluyen otros puntos. Se tiene que la primera componente es de tamaño y recoge prácticamente toda la información. La clasificación deja 5 grupos claramente diferenciados y geográficamente distribuidos en zonas. Se clasifica las provincias en tres grupos según sea su mortalidad frente a la nacional más favorable o menos favorable y un tercer grupo con las provincias que no queda claro y a las cuales se aplica un análisis de cluster que determina tres grupos según la mortalidad sea casi siempre favorable o casi siempre desfavorable y otro grupo indefinido. Así, se tiene que se repite la distribución geográfica de la mortalidad de las provincias en España.

El primero y principal objetivo del análisis de la mortalidad es proporcionar herramientas para la estimación y proyección de la población. El estudio descriptivo de este capítulo consolida la base para plantear modelos relacionales entre el nacional y el provincial. Los modelos relacionales, con larga tradición, explican la mortalidad de las provincias en función de la mortalidad del total nacional (Brass, 1971) y permiten

aprovechar la estabilidad de las poblaciones de tamaño grande a las que incorporan una componente diferencial de la población de pequeño tamaño.

En el capítulo siguiente se revisan los distintos procedimientos utilizados históricamente para la estimación y proyección de los parámetros provinciales y avanzar más allá de la estimación puntual apostando por la estimación por intervalos y de esta manera se consigue adicionalmente, proporcionar una medida de la bondad de la predicción.

Capítulo 4

Estudio sobre la Proyección de la mortalidad en poblaciones de pequeño tamaño

- 4.1. Introducción
- 4.2. Fundamentos teóricos
- 4.3. Aproximación histórica
- 4.4. Proyección del riesgo de muerte provincial
- 4.5. Proyección de la esperanza de vida provincial
- 4.6. Resultados y conclusiones

4.1. Introducción

En el capítulo 3 de este trabajo se pone de manifiesto que la incidencia de la mortalidad en poblaciones grandes presenta un comportamiento estable. Esta estabilidad facilita su estimación y proyección a corto y largo plazo.

Además del tamaño de la población, distintos factores afectan a las oscilaciones de las series. Tienen una importante influencia sobre las oscilaciones el tamaño de la variable e incluso la elección de la función objeto de estudio. A diferencia de las poblaciones grandes, las poblaciones de pequeño tamaño presentan una fluctuación más acusada que complica su estimación y proyección.

El objetivo del presente capítulo es establecer un criterio para la estimación por intervalos en poblaciones de pequeño tamaño. Frente al tradicional enfoque clásico de la estimación puntual, se pretende diseñar una estrategia que responda a la necesidad de, además de calcular un valor de predicción, proporcionar una medida de la bondad de la predicción. En esta línea juega un importante papel la elección de la función biométrica que, de acuerdo con sus propiedades, mejor responda a las necesidades de la predicción.

Se aplicará un estudio sobre la incidencia de la mortalidad en las provincias de España, prestando una particular atención al tratamiento que requiere la controvertida y creciente existencia de riesgos de muerte nulos, como se ha mencionado en el capítulo 2, pasando luego a tratar la alternativa que ofrece la esperanza de vida con sus ventajosas características.

Este capítulo se desarrolla en 6 apartados.

- En el segundo apartado, se presentan los fundamentos teóricos que gobiernan la incidencia de la mortalidad provincial a través de las funciones biométricas de la tabla de mortalidad.
- El tercero consiste en una breve revisión histórica hasta la fecha de distintas estrategias que encontramos en la literatura para enfrentar el reto de la estimación y predicción de la mortalidad en las provincias de España, en particular la metodología actual del INE.

- En el cuarto apartado se afronta el establecimiento de un protocolo de actuación para la estimación y proyección del riesgo de muerte con toda la casuística que conlleva esta función en las poblaciones provinciales.
- En el apartado quinto se presenta un protocolo de actuación para la estimación y proyección de la esperanza de vida en las provincias que incluye un intervalo de confianza para la predicción.
- Se termina con un sexto apartado donde se recogen los resultados y conclusiones del capítulo y se comparan las alternativas de estimación y proyección del riesgo de muerte frente a la esperanza de vida.

4.2 Fundamentos teóricos

El riesgo de muerte en una población, tanto a nivel nacional $q(t, x)$ como provincial $q(t, x, h)$, observado cada año constituyen una serie de tiempo, realización de un proceso estocástico.

Estos procesos estocásticos se definen de la siguiente manera:

Sea \mathfrak{R} conjunto de los números reales. Para cada edad $x \in \mathfrak{R}$ con $x \geq 0$ y cada instante $t \in \mathfrak{R}$ con $t \geq 0$ se define el proceso estocástico $\{q(t, x) : t \in \mathfrak{R}\}$ donde $q(t, x)$ es la probabilidad de que un individuo de la población objeto de estudio, de edad x años en el momento t , muera antes del momento $t+1$.

Se trata de una extensión al espacio continuo para $t \in \mathfrak{R}$ y $x \geq 0$ de la definición de $q(t, x)$ descrita en la metodología del segundo capítulo para cualquier entero t y cualquier entero $x \geq 0$.

Para cada año t y edad x $q(t, x)$ es una variable aleatoria cuyo soporte es el intervalo $[0, 1]$. El intervalo incluye los extremos porque se trata de una probabilidad que puede tomar tanto el valor 0 como el 1.

Análogamente se definen los procesos estocásticos $\{m(t, x) : t \in \mathfrak{R}\}$ donde $m(t, x)$ es la tasa de mortalidad; $\{a(t, x) : t \in \mathfrak{R}\}$ donde $a(t, x)$ es el tiempo medio vivido por los que se mueren antes de pasar un año; $\{l(t, x) : t \in \mathfrak{R}\}$ donde $l(t, x)$ es el

número de supervivientes; $\{d(t, x) : t \in \mathfrak{R}\}$ donde $d(t, x)$ es el número de defunciones; $\{L(t, x) : t \in \mathfrak{R}\}$ donde $L(t, x)$ es la población estacionaria y $\{e(t, x) : t \in \mathfrak{R}\}$ donde $e(t, x)$ es la esperanza de vida.

Para cada (t, x) fijo $m(t, x)$ es una variable aleatoria cuyo soporte es el conjunto de números reales mayores o iguales que cero; el soporte de la variable $a(t, x)$ es el intervalo $[0, 1]$; Para la variable $l(t, x)$ es el intervalo $[0, l(t, x - 1)]$; para la variable $d(t, x)$ es $[0, l(t, x)]$; para la variable $L(t, x)$ es $[l(t, x + 1), l(t, x)]$ y para la variable $e(t, x)$ es el conjunto de números reales positivos.

Para cada edad de 0 a $w - 1$ años la tabla de mortalidad de cada año incluye una realización de cada proceso estocástico definido. Hay que recordar que w es la edad de cierre de la tabla, 100 para el total nacional y 95 para las provincias excepto Ceuta y Melilla que cierran a 90.

Para el caso particular del último grupo de edad la definición de los procesos estocásticos de las funciones de la tabla de mortalidad es similar teniendo en cuenta que la región del diagrama de Lexis asociada a los procesos es un intervalo abierto.

Dado $w \in \mathfrak{R}$ con $w \geq 0$ y para cada instante $t \in \mathfrak{R}$ con $t \geq 0$ se define el proceso estocástico $\{q(t, w) : t \in \mathfrak{R}\}$ donde $q(t, w)$ es la probabilidad de que un individuo de la población objeto de estudio, de edad w años se muera en el momento t . Esta probabilidad por definición es 1. Es decir se trata de un proceso estocástico degenerado.

Análogamente, se definen los procesos estocásticos $\{m(t, w) : t \in \mathfrak{R}\}$ donde $m(t, w)$ es la tasa de mortalidad en la región abierta del diagrama de Lexis determinada por la edad w y los instantes t y $t+1$; $\{a(t, w) : t \in \mathfrak{R}\}$ donde $a(t, w)$ es el tiempo medio vivido por los fallecen en la región abierta y por definición es $1/m(t, w)$; $\{l(t, w) : t \in \mathfrak{R}\}$ donde $l(t, w)$ es el número de supervivientes que coincide con $\{d(t, w) : t \in \mathfrak{R}\}$ donde $d(t, w)$ es el número de defunciones; $\{L(t, w) : t \in \mathfrak{R}\}$ donde $L(t, w)$ es la población estacionaria en la región abierta y $\{e(t, w) : t \in \mathfrak{R}\}$ donde $e(t, w)$ es la esperanza de vida que coincide con el tiempo medio vivido a la edad w .

Para cada $t \geq 0$ fijo $q(t, w)$ es una variable aleatoria que vale 1 siempre. $m(t, w)$ es una variable aleatoria cuyo soporte es el conjunto de números reales mayores o iguales que cero; el soporte de la variable $a(t, w)$ es el conjunto de números reales mayores o iguales que cero. Para la variable $l(t, w)$ es el intervalo $[0, l(w-1)]$; para la variable $d(t, w)$ es $l(t, w)$; para la variable $L(t, w)$ y la variable $e(t, w)$ es el conjunto de números reales positivos.

En cuanto a la fundamentación muestral, la estimación de la tasa de mortalidad y el riesgo de muerte para la edad x esta condicionada por el tamaño de la población a esa edad. En cambio, la estimación de la esperanza de vida a la edad x se construye a partir de la esperanza de vida a edades superiores. Es decir, el conjunto de datos considerado se extiende a la población de x y más años. De esta forma se amplía el tamaño de la muestra y aumenta la consistencia de la estimación, que será tanto mayor cuanto menor sea la edad, alcanzando su máximo con la esperanza de vida al nacimiento, en cuya estimación se incluyen todos los individuos de la población objeto de estudio.

Análogamente también las funciones de supervivientes $l(x)$ y la población estacionaria $L(x)$ tienen una base muestral más grande cuanto mayor sea la edad y por tanto serán funciones más estables.

Por otra parte, una importante cuestión a tener en cuenta es que a partir de cualquier función de las tablas de mortalidad se puede reconstruir el resto de la tabla, dada la función $a(x)$, tiempo vivido el año de defunción. Es decir a partir del riesgo de muerte se puede obtener la esperanza de vida y viceversa. Así pues a la hora de proyectar la mortalidad podemos aprovechar la función que mejor satisfaga las necesidades de la estimación y proyección.

4.3 Aproximación histórica

A lo largo de la Historia el interés por conocer la evolución de la población se ha traducido en una abundante literatura. El lector interesado en ampliar sus conocimientos metodológicos sobre los modelos existentes para la predicción de la mortalidad puede consultar las monografías de Girosi y King (2007) y Camarda (2008).

Se distinguen dos familias de métodos dependiendo de si hay o no variables covariantes, i.e. variables que proporcionan información exógena para mejorar la predicción de la mortalidad. Los resultados de este primer bloque de modelos son discutibles porque implican en muchos casos la predicción de otras variables como el consumo de tabaco, el Producto Interior Bruto o nivel educativo, tan complicadas o más de predecir que la propia mortalidad.

En cuanto a los modelos sin covariantes, se basan en la suposición de que toda la información sobre el futuro esta contenida en la incidencia de la mortalidad pasada. Los impactos exógenos como los descubrimientos de nuevas tecnologías médicas, crisis económicas o innovaciones en la salud pública se ignoran, mientras que los ciclos epidemiológicos predecibles debidos a respuestas biológicas o de comportamiento reflejadas en la mortalidad pasada se incluyen (ver Gutterman y Vanderhoof, 1998).

Entre los modelos de este bloque tenemos en primer lugar los paramétricos. Estos modelos tienen larga tradición en la Demografía clásica y se basan en la pura extrapolación de la función biométrica elegida, riesgo de muerte, tasa de mortalidad, supervivientes, esperanza de vida u otra. La idea es resumir la información de la mortalidad recogida por el indicador en una función parsimoniosa pero lo suficientemente flexible para recoger la evolución en el tiempo. La función es de la forma general:

$$f(t, x) = g(x, b(x), c(t)) + \varepsilon(t, x),$$

donde $g(x, b(x), c(t))$ es la parte determinista, con $b(x)$ y $c(t)$ funciones de la edad y del tiempo y $\varepsilon(t, x)$ la parte aleatoria.

Gompertz (1825) fue el primer demógrafo que observó un crecimiento exponencial de la mortalidad a partir de los 20 años y especificó el modelo:

$$\ln q(t, x) = a + b * x + \varepsilon(t, x).$$

Desde entonces, se han hecho docenas de propuestas para la función f . Algunos modelos clásicos de la literatura son los de Keyfitz (1968, 1982) o Tabeau, Berg y Heathcote (2001) hasta llegar al modelo de Heligman & Polard (1980) con 8 parámetros constantes para cada año.

Otra corriente de modelos sigue una aproximación no paramétrica a través del Análisis de Componentes Principales. Su exponente más conocido es el denominado modelo Lee-Carter (Lee y Carter, 1992). La literatura ha usado una gran variedad de técnicas para intentar modelizar los patrones de la mortalidad pero ninguno es completamente satisfactorio. Aún así, el objetivo de reducir los datos a sus elementos nucleares y producir predicciones que incluyan los rasgos conocidos de los datos es crítico y todos los países a través de sus Oficinas de Estadística hacen estimaciones y proyecciones.

En España la primera aproximación a la proyección de la población es un folleto publicado por el INE, INE (1941), donde se estima el tamaño de la población ajustando a una curva logística por el método de Verhulst. El siguiente ejercicio de proyección, INE (1956), vuelve a ajustar una curva logística por el método anterior y prueba el método de Pearl con el que obtiene un mejor ajuste. Para finales del siglo XX se predecía una población de 39 millones de residentes frente a la cifra de 40.847.371 residentes que da el censo de 2001 y con una límite de 44 millones.

Hasta entonces solo se proyectó el tamaño de la población sin atender a la estructura de la población por sexo y edad ni su desagregación territorial. La mejora de la calidad de la información y la potencia de las herramientas estadísticas permitió aplicar el método de las componentes que sigue el siguiente esquema: a partir de la

población de un censo se establecen hipótesis sobre la evolución que van a experimentar en el futuro las tres componentes demográficas básicas -la mortalidad, la fecundidad y las migraciones- que determinan su crecimiento y su estructura por edades. En la monografía del INE (1976) se recoge un análisis de la evolución histórica de la población y se aplica el método de las componentes para las proyecciones 1970-1995 en base al censo de 1970. La hipótesis sobre mortalidad fue considerarla constante en base la dificultad de mantener las espectaculares ganancias registradas con anterioridad. Se tomo como incidencia de la mortalidad la calculada en las tablas de mortalidad (INE, 1978) con periodo de referencia de cuatro años, desde 1969 a 1972. De acuerdo con INE (1995) para 1980 se proyectó una esperanza de vida de 70,5 años y 76,4 años, para varones y mujeres respectivamente, superados por la realidad en 2 y 2,2 años respectivamente. En INE (1981) se revisaron las proyecciones del periodo 1978-1995 con población de partida del padrón de 1975, como continuación del estudio del INE (1980) sobre la evolución de la población española en el periodo 1961-1978.

El siguiente ejercicio, aprovechando como base el censo de 1981, es la proyección con los resultados del conjunto nacional, INE (1987) y por Comunidades Autónomas, INE (1988c). Se supuso la mortalidad constante en todo el periodo considerado.

La siguiente proyección INE (1995) es la referida al periodo 1991-2020 a partir de la reconstrucción de los datos intercensales del periodo 1981-1990. Se calculan proyecciones referidas al principio y mitad de año con un periodo de treinta años para el total nacional y quince para las comunidades autónomas y las provincias. Se obtienen poblaciones, por sexo y edades simples, para el conjunto nacional y las comunidades autónomas, hasta 100 años y más y 85 años y más, respectivamente (Cantalapiedra, 2001).

Las proyecciones de la población de España calculadas a partir del Censo de población de 1991 sigue el principio de extrapolar las tendencias por edades simples en intervalos temporales quinquenales, mediante la formula:

$$m(t+5, x) = m(t, x) \cdot \gamma(x), \quad (4.1)$$

donde $m(t, x)$ representa la tasa de mortalidad a la edad x en el año t , y $\gamma(x)$ es el coeficiente de mejora anual de la mortalidad a la edad x . A continuación se suaviza la serie de los coeficientes de mejora observados. En este cometido se utiliza una interpolación por segmentos a partir de ajustes por mínimos cuadrados ordinarios en distintos tramos por edades elegidos bajo criterios discrecionales. La edad 0 se trata separadamente observando la evolución de la tendencia a lo largo de los años.

Para la proyección de las tablas de mortalidad de las Comunidades Autónomas se calcula un índice que muestra la importancia de la mortalidad a una cierta edad con respecto a la correspondiente del total nacional según la fórmula:

$$I(t, x, CCAA) = \frac{m(t, x, CCAA)}{m(t, x)}, \quad (4.2)$$

donde $m(t, x, CCAA)$ representa la tasa de mortalidad a la edad x en la Comunidad Autónoma y $m(t, x)$ es la tasa de mortalidad a la edad x en el total nacional. La proyección de mortalidad de cada comunidad se calcula en función de la tabla nacional y dicho índice. La tasa de mortalidad para cada Comunidad Autónoma se calcula para un periodo de 10 años en base a las defunciones adecuadamente clasificadas del movimiento natural de la población y los censos de 1981 y 1991.

En cuanto a la mortalidad infantil se siguió la hipótesis de convergencia de las comunidades autónomas al conjunto nacional.

En 2003 se publican las proyecciones de mortalidad, INE (2003), a partir de los datos observados hasta 2001. Para el total nacional se proyecta el periodo 2002 a 2032. La función proyectada es el riesgo de muerte por sexo y edad a nivel nacional a partir de los coeficientes de variación de mortalidad definidos como:

$$r(t, x) = \frac{q(t+1, x)}{q(t, x)} \quad (4.3)$$

para cada año desde 1991 al 2000. Estos coeficientes se suavizan anualmente por tramos de edades y se proyecta su tendencia.

En el caso de las comunidades autónomas el periodo de proyección es desde 2002 a 2017. La función estimada es el riesgo de muerte por sexo y edades simples utilizando unos índices definidos como:

$$I(t, x, CCAA) = \frac{q(t, x, CCAA)}{q(t, x)}. \quad (4.4)$$

A partir de la serie de índices en el periodo se extrapola la tendencia y con los valores para cada año se tiene que el riesgo de muerte proyectado de cada comunidad autónoma es:

$$\hat{q}(t+k, x, CCAA) = \hat{I}_{t+k, x, CCAA} \cdot \hat{q}(t+k, x). \quad (4.5)$$

En el caso de la mortalidad provincial se iguala a la de la correspondiente comunidad autónoma.

Actualmente el INE elabora, a nivel nacional, estimaciones a tiempo real, y proyecciones tanto a corto cada año, i.e. con un horizonte de 10 años, como a largo plazo, i.e. con un horizonte de 40 años cada tres años. Las proyecciones a corto plazo se realizan todos los años a finales de año, el último ejercicio se publicó en noviembre, INE (2013b). La metodología es la misma desde 2008 con ligeros cambios en los periodos de observación o en suavizado. A nivel provincial solo se elaboran estimaciones a tiempo real y proyecciones a corto. Esta actitud se justifica ante la dificultad de garantizar un mínimo de acuracidad. Si a nivel nacional ya es difícil predecir, aún cuando la mortalidad tiene un comportamiento razonablemente estable, tratar de proporcionar datos a nivel provincial a más de 10 años vista es poco menos que especular.

Esta metodología para la estimación y proyección de la mortalidad en las provincias toma como punto de partida la estimación y proyección de la mortalidad para el total nacional.

La metodología de estimación de la incidencia de la mortalidad en España se lleva a cabo a partir de una extrapolación de las tendencias observadas de los riesgos o probabilidades de muerte a cada edad, de acuerdo a una modelización exponencial negativa de las trayectorias suavizadas de las mismas en función del tiempo. Se desarrolla en las etapas que a continuación se describen:

1. En primer lugar, se suaviza la serie temporal, que abarca los últimos diez años, compuesta por las probabilidades de muerte $q(t, x)$ en cada año t y edad x observadas, deducidas de las tablas de mortalidad anuales del INE hasta el último año para el que se dispone de resultados definitivos de la misma, y una estimación adelantada de dichas probabilidades referidas a periodos de doce meses consecutivos para el periodo más reciente procedente de los últimos resultados disponibles de las Estimaciones Mensuales de Coyuntura Demográfica (resultados avanzados de los Indicadores Demográficos Básicos). Dicho proceso se lleva a cabo a partir de un doble suavizado de medias móviles de orden cinco (cinco edades consecutivas).

2. Se ajusta al siguiente modelo no lineal del tiempo (años) la serie suavizada de probabilidades de muerte para las edades de 0 a 99:

$$\ln q(t, x) = \alpha(x) + \beta(x) \cdot t. \quad (4.6)$$

Los parámetros $\alpha(x)$ y $\beta(x)$ de cada uno de estos modelos se estiman por Mínimos Cuadrados Ordinarios aplicados a los modelos lineales, obteniéndose los valores $\hat{\alpha}(x)$ y $\hat{\beta}(x)$, respectivamente.

3. Se suaviza la serie por edad estimada del parámetro $\hat{\beta}(x)$ mediante un doble proceso de suavizado de medias móviles de orden cinco para evitar divergencias indeseables en las evoluciones de las mejoras en mortalidad de edades consecutivas.

5. Estos parámetros $\hat{\alpha}(x)$ sin suavizar y $\hat{\beta}(x)$ suavizado son los empleados para predecir, según dicha modelización, la probabilidad de muerte, para cada edad, en el año correspondiente:

$$\hat{q}(t, x) = \exp \{ \hat{\alpha}(x) + \hat{\beta}(x) \cdot t \} \quad (4.7)$$

6. Finalmente, a partir de las probabilidades de muerte estimadas para el año en curso, se deriva cada una de las funciones biométricas de la tabla de mortalidad completa. Así, partiendo de una generación ficticia de $l(t,0) = 100.000$ individuos se obtienen:

- Las defunciones de edad x:

$$d(t, x) = l(t, x) \cdot q(t, x) \quad (4.8)$$

- Los años vividos con edad x:

$$L(t, x) = l(t, x + 1) + a(t, x) \cdot d(t, x), \quad (4.9)$$

donde $a(t, x)$ es el tiempo promedio de vida con edad de los fallecidos con dicha edad, estimado mediante Mínimos Cuadrados Ordinarios según un modelo log-lineal de los valores anuales observados para los últimos nueve años y recogidos en las defunciones registradas en la Estadística del Movimiento Natural de la Población sobre el tiempo t, es decir:

$a(t, x) = \gamma(x) + \delta(x) \ln t$, con $t=1, \dots, 9$. Siendo 1 el primer año observado, etc. para que el modelo no degenera en un ajuste lineal.

- Las tasas de mortalidad por generación:

$$m(t, x) = \frac{L(t, x) - L(t, x + 1)}{\frac{1}{2}(L(t, x) + L(t, x + 1))} \quad (4.10)$$

- El tiempo vivido desde la edad:

$$T(t, x) = T(t, x + 1) + L(t, x) \quad (4.11)$$

- La esperanza de vida a cada edad:

$$e(t, x) = \frac{T(t, x)}{l(t, x)}. \quad (4.12)$$

Una vez calculada la estimación o proyección nacional del año objeto de interés, se calcula el valor a nivel provincial usando una metodología relacional que liga la estimación provincial a la establecida para el nivel nacional, siguiendo el método de los logits de Brass (Brass, 1975).

El procedimiento de estimación se desarrolla en los siguientes pasos:

1. Se parte del cálculo de la función de supervivientes por edad cumplida x , de las tablas de mortalidad¹ anuales completas de cada provincia y de España, las cuales denotamos por $l(t, x, h)$ y $l(t, x)$ para cada año t , respectivamente.

2. Se realiza la siguiente transformación logística a las series anuales de supervivientes de cada sexo calculadas para cada provincia y para España:

$$\text{Logit } l(t, x, h) = \frac{1}{2} \ln \left(\frac{l(t, 0, h) - l(t, x, h)}{l(t, x, h)} \right) \quad (4.13)$$

$$\text{Logit } l(t, x) = \frac{1}{2} \ln \left(\frac{l(t, 0) - l(t, x)}{l(t, x)} \right) \quad (4.14)$$

La serie de supervivientes de cada provincia y la del total nacional tienen la propiedad de que la relación entre los valores transformados de ambos es aproximadamente lineal, de manera que, para cada año t , es susceptible de ser modelizada mediante una recta de regresión:

$$\text{Logit } l(t, x, h) = \alpha(t, h) + \beta(t, h) \times \text{Logit } l(t, x) \quad (4.15)$$

En el ajuste de dichos modelos se emplean únicamente los valores de las series que van de los 40 a los 95 años (en Ceuta y Melilla hasta los 90). Diversas razones justifican dicho proceder:

¹ INE, Tablas de mortalidad. 1991-2011 [www.ine.es]

1. En primer lugar, porque la función de supervivientes es una función que acumula la incidencia de la mortalidad sobre la población. La función de supervivientes a la edad x , $l(x)$ se construye a partir de la incidencia de la mortalidad en la población con x y menos años. Los efectos aleatorios, a los que está expuesta la función debido a la coyuntura y a los sucesos raros, más cuanto menor sea el tamaño de la población, se amortiguan al acumularse años y con ello, población observada.

2. Es necesario observar que los riesgos de morir en la infancia, la adolescencia y la primera adultez son más vulnerables a las oscilaciones que en edades posteriores. Es esa importante aleatoriedad en la mayoría de los territorios la que provoca que la función de supervivencia presente fluctuaciones, que podrían repercutir indeseablemente sobre los valores de los parámetros de la recta de regresión, si se utilizaran todas las edades para su cálculo.

3. Por otro lado, el progresivo desplazamiento de la fuerza de la mortalidad a edades cada vez más avanzadas, ha provocado una paulatina pérdida del protagonismo de la infancia y de la adolescencia en la explicación de los diferenciales territoriales de vida media entre las provincias y España, proceso que ha sido más acentuado en las mujeres, ya que se encuentran en un estadio más avanzado de la transición epidemiológica. Como excepción, puede señalarse la desigual incidencia territorial del repunte de la mortalidad en los adultos más jóvenes, especialmente en los hombres, en la década de los ochenta y buena parte de los años noventa. No obstante, la reciente trayectoria de descenso de la mortalidad en esas edades ha provocado que vuelvan a perder peso en la explicación de los diferenciales espaciales de vida media.

4. En las poblaciones que gozan de bajos niveles de mortalidad, el impacto de las hipótesis de mortalidad se concentra en las edades maduras y, sobre todo, en las avanzadas. Si bien en términos del valor de la esperanza de vida al nacer no son desdeñables los riesgos de morir en la primera mitad de la vida, su impacto sobre los efectivos poblacionales a la postre estimados o proyectados es menor que en las edades maduras y avanzadas. Por tanto, en este proceso de estimación es más

relevante la previsión del comportamiento de los riesgos de morir actuales en la madurez y en la ancianidad.

Los parámetros anuales $\alpha(t, h)$ y $\beta(t, h)$ describen el nivel y la estructura de mortalidad de cada provincia en relación con el total nacional en cada año, de forma que un valor negativo de $\alpha(t, h)$ indica un comportamiento general de la mortalidad más favorable en la correspondiente provincia que en el conjunto de España en dicho año, y viceversa. Un valor de $\beta(t, h)$ superior a uno indicará que la incidencia de la mortalidad en tal provincia es más favorable en las primeras etapas de la vida que en las edades avanzadas en relación con la del conjunto nacional en ese año, y viceversa. Dichos parámetros se estiman por el método de Mínimos Cuadrados Ordinarios, obteniéndose los valores $\alpha(t, h)$ y $\beta(t, h)$ para cada provincia y para cada año.

Finalmente, se establece una relación log-lineal para la evolución de tales parámetros en cada provincia en función del tiempo que, estimada por Mínimos Cuadrados Ordinarios haciendo uso de los datos observados para los nueve últimos años disponibles, permite derivar la estimación de los mismos para el año en curso²:

$$\alpha(t, h) = \lambda(h) + \rho(h) \cdot \ln t \quad (4.16)$$

$$\beta(t, h) = \pi(h) + \vartheta(h) \cdot \ln t \quad (4.17)$$

Para $t=1, \dots, 9$ donde 1 es el primer año observado, etc. para que el modelo no degenerare en un ajuste lineal.

² Desde el segundo trimestre hasta el último de 2010, se eligió como estimación de estos parámetros para el año en curso, los derivados del ajuste del modelo log-lineal con los datos desde 1991 hasta el último año con resultados definitivos de las Estadísticas del Movimiento Natural de la Población disponibles. Anteriormente se eligió como estimación los derivados de funciones de supervivientes promedios de las calculadas para los cuatro últimos años.

4.4 Proyección del riesgo de muerte provincial

Tradicionalmente, las proyecciones de la mortalidad provincial proporcionan un valor para el riesgo de muerte de una edad determinada sin tener en cuenta el importante componente de fluctuaciones que se producen en pequeñas áreas. Se hace necesario proporcionar, además de la estimación puntual una estimación por intervalos como medida de la bondad de esa estimación puntual para un nivel de confianza dado.

En este sentido, es clave la elección de la variable que mide la incidencia de la mortalidad. Se pueden distinguir dos aproximaciones al problema. Una primera aproximación consiste en utilizar de entre todas las funciones biométricas de la tabla de mortalidad, variables que conservan la independencia por edades como la tasa o el riesgo. Es el criterio utilizado tradicionalmente en la proyección y estimación en poblaciones grandes. De la proyección o estimación del riesgo o la tasa de muerte se derivan el resto de las funciones biométricas de la tabla de mortalidad. Por construcción, estas funciones están expuestas a los factores aleatorios intrínsecos de las poblaciones de pequeño tamaño.

La otra aproximación consiste en utilizar variables como los supervivientes, el tiempo que queda por vivir o la esperanza de vida, que son dependientes por edad y que amortiguan el efecto de los factores aleatorios de las poblaciones pequeñas, al calcularse a partir de muestras que incluyen más población. Esta alternativa con la función de supervivientes es la que se utiliza actualmente al aplicar el logit de Brass. A continuación, se describe un protocolo de actuación para el cálculo del intervalo de confianza para la estimación del riesgo de muerte provincial. Merece especial atención, en este caso, el tratamiento de los ceros, escollo que se trata de salvar con el siguiente protocolo:

Dado el riesgo de muerte del total nacional $q(t, x)$, para cada provincia h se define su *diferencial provincial* como el proceso estocástico de los restos, $r_q(t, x, h) = q(t, x, h) - q(t, x)$ que tiene como soporte $[-q(t, x), 1 - q(t, x)]$

Como $q(t, x, h) = q(t, x) + rq(t, x, h)$ la estimación de $q(t, x, h)$ se construye como suma de la estimación del riesgo para el total nacional y el intervalo de confianza proporcionado por la estimación de estos diferenciales.

Distintas razones llevan a elegir $rq(t, x, h)$ en vez de $q(t, x, h)$ como objeto de estudio en la mortalidad provincial. En primer lugar, elimina la influencia coyuntural de la mortalidad. Cuando las circunstancias climatológicas, epidémicas, etc. den lugar a un aumento de la mortalidad el riesgo de muerte aumentará tanto a nivel nacional como provincial. No así a este diferencial provincial, imputable únicamente al especial comportamiento de la provincia respecto del total nacional. En segundo lugar evita los ceros del riesgo de muerte que complican la estimación del modelo ya que la especificación tiene que basarse en distribuciones mixtas al aparecer valores discretos con probabilidades positivas.

Para modelizar el diferencial de la provincia h con relación al total nacional, se especifica el siguiente modelo:

$$rq(t, x, h) = a(x, h) + b(x, h) \cdot t + \varepsilon(t, x, h) \text{ para } t=1,2,\dots,18 \quad (4.18)$$

La ordenada en el origen $a(x, h)$ es un parámetro particularmente sensible a la localización de la nube de puntos respecto al eje de ordenadas. Para limitar la influencia de la pendiente en el ajuste de este coeficiente se hace un cambio de localización de tal forma que $t=1$ representa el año 1991.

A continuación se desarrollan los pasos para la obtención de un intervalo de confianza. Para un estudio más exhaustivo del modelo lineal puede consultarse el manual de referencia de Myers (1990).

A partir de la suma de cuadrados de los residuos de la tabla del Análisis de la varianza se estima la varianza σ^2 mediante el estadístico :

$$S^2 = \frac{\|rq(t, x, h) - r\hat{q}(t, x, h)\|^2}{n - 2} \quad (4.19)$$

Sea $t_{1-\alpha/2}$ tal que $P\left[|T| < t_{1-\alpha/2}\right] = 1 - \alpha$ siendo T una variable aleatoria con distribución t de Student con $n-2$ grados de libertad se tiene que:

$$\left|rq(t, x, h) - \hat{a}(x, h) - \hat{b}(x, h) \cdot t\right| \leq \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{xx}}} \cdot S \cdot t_{1-\alpha/2} \quad (4.20)$$

El modelo proporciona el siguiente intervalo de confianza para el resto:

$$rq(t_0, x, h) \in \left[\hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{xx}}} \cdot S \cdot t_{1-\alpha/2} \right] \quad (4.21)$$

Por tanto, de acuerdo con el método de estimación por intervalos se tiene que para un nivel de confianza $1 - \alpha$, la estimación de $q(t_0, x, h)$ tiene como intervalo de confianza $[Inf(t_0, x, h), Sup(t_0, x, h)]$ donde:

$$Inf(t_0, x, h) = \max\{0, q(t_0, x) + \hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 - \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{xx}}} \cdot S \cdot t_{1-\alpha/2}\} \quad (4.22)$$

$$Sup(t_0, x, h) = \min\{1, q(t_0, x) + \hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 + \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{xx}}} \cdot S \cdot t_{1-\alpha/2}\} \quad (4.23)$$

Los extremos se definen así porque el soporte del riesgo está en el intervalo acotado por el 0 y el 1, ambos inclusive. Por tanto en el caso de que el intervalo incluya los números negativos o los superiores a 1 se le asigna el extremo correspondiente del intervalo.

Hay un punto especialmente conflictivo en las proyecciones tradicionales en pequeñas áreas. De acuerdo con la metodología, el valor del riesgo en ningún caso será cero. Este hecho contradice la observación de la serie. El riesgo observado puede ser cero y esto ocurre en bastantes casos. Por tanto, es un escenario que debe contemplar la predicción. Por supuesto nadie pretende decir que la predicción del riesgo sea cero. Se trata de que la predicción sea un cierto valor cuyo intervalo de confianza pueda incluir al cero. Es decir, se contemple la posibilidad de riesgo nulo.

En el caso de que exista al menos un valor nulo en la serie provincial de riesgo de muerte o aún siendo todos distintos de cero el análisis de residuales rechaza la

normalidad para los errores muestrales de la regresión lineal, los restos se ajustan al modelo probabilístico cuya función de distribución mixta se define como:

$$F(z) = \begin{cases} 0 & \text{si } z < a \\ p & \text{si } z = a \\ p + (1-p) \cdot \Phi\left(\frac{z-\mu}{\sigma}\right) & \text{si } z \in (a, 1-a) \\ 1 & \text{si } 1-a \geq 1 \end{cases} \quad (4.24)$$

donde

- $\Phi(\cdot)$ es la función de distribución de una normal tipificada.
- a es la cota inferior del soporte determinado por el total nacional y toma como valor el riesgo de muerte nacional con signo negativo.
- p es la probabilidad del punto a
- μ y σ son los parámetros de localización y escala para la parte continua de la distribución mixta.

Bajo la hipótesis de que los diferenciales provinciales son variables aleatorias independientes e igualmente distribuidas, y dado que la distribución de $rq(t, x, h)$ tiene como soporte $[-q(t, x), 1 - q(t, x)]$, el procedimiento de estimación de los distintos parámetros se desarrolla en los siguientes pasos:

1. El valor estimado de $rq(t, x, h)$ en el horizonte $k=1$ tiene como soporte el intervalo predicho $[-\hat{q}(t+1, x), 1 - \hat{q}(t+1, x)]$. Se estima el parámetro a por el valor predicho $-\hat{q}(t+1, x)$.

2. Una vez estimado el parámetro a del modelo probabilístico todas las observaciones $rq(t, x, h)$ que se salgan del intervalo predicho se imputan al inferior o al superior dependiendo de si quedan por arriba o por abajo.

3. El parámetro p se estima como la frecuencia relativa de la cota inferior del soporte.

4. Los parámetros μ y σ se estiman por el método de los momentos. El primer momento poblacional es:

$$E[X] = a \cdot p + (1-p) \cdot \left[\mu \cdot \left(1 - \Phi\left(\frac{a-\mu}{\sigma}\right)\right) + \frac{\sigma}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{a-\mu}{\sigma}\right)^2\right\} \right] \quad (4.25)$$

Y el momento poblacional de orden dos es:

$$E[X^2] = a^2 \cdot p + (1 - p) \left[\mu^2 (1 - \Phi(\frac{a - \mu}{\sigma})) + r(a, \mu, \sigma) \right], \quad (4.26)$$

con:

$$r(a, \mu, \sigma) = \frac{\sigma^2}{\sqrt{2\pi}} \left(\left(\frac{a - \mu}{\sigma} \right) \exp \left\{ -\frac{1}{2} \left(\frac{a - \mu}{\sigma} \right)^2 \right\} + \sqrt{2\pi} (1 - \Phi(\frac{a - \mu}{\sigma})) \right) + \frac{2\mu\sigma}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{a - \mu}{\sigma} \right)^2 \right\}$$

La proyección para horizontes de k mayores que 1 sigue el mismo procedimiento. En primer lugar se estima a con la proyección del valor nacional para el horizonte deseado y una vez obtenido este valor se siguen los siguientes pasos descritos en el proceso para un horizonte k igual a 1 para la estimación de p , μ y σ .

La proyección no tiene en cuenta el horizonte k para los parámetros p , μ y σ con lo que se pierde fiabilidad con el paso del tiempo ya que el modelo no es capaz de captar la pérdida de precisión y el aumento de la amplitud del intervalo de confianza.

La aparición de valores nulos para el riesgo provincial es cada vez más frecuente por lo que se hace necesario recurrir al caso particular del modelo (4.24) cada vez más.

Como se puede apreciar la estrategia de proyectar la mortalidad utilizando el riesgo de muerte tiene importantes inconvenientes. Son los ya conocidos que se derivan del tamaño de la muestra. A saber, la poca estabilidad de las series de tiempo a proyectar y aparición de riesgos nulos observados principalmente.

Una interesante línea de investigación por explorar es utilizar como función a proyectar en vez del riesgo de mortalidad, la esperanza de vida que por definición incorpora un mayor tamaño de la muestra para estimar. Esta es la vía que se presenta a continuación.

4.5 Proyección de la esperanza de vida provincial

Frente a las complejas características del riesgo de mortalidad provincial, la esperanza de vida constituye una variable más consistente y estable a lo largo del tiempo, incluso a nivel provincial.

A continuación se describe un protocolo de actuación para la estimación del intervalo de confianza para la esperanza de vida provincial.

Análogamente al proceso empleado para el riesgo de muerte provincial, dada la esperanza de vida del total nacional $e(t, x)$ se define el proceso estocástico de los *diferenciales*:

$$re(t, x, h) = e(t, x, h) - e(t, x), \quad (4.27)$$

que tiene como soporte el conjunto de los números reales. Despejando,

$$e(t, x, h) = e(t, x) + re(t, x, h), \quad (4.28)$$

luego la esperanza de vida de la provincia h , $e(t, x, h)$ se construye como suma de la esperanza de vida para el total nacional y su correspondiente diferencial.

Distintas razones llevan a definir la esperanza provincial como la agregación de dos sumandos. En primer lugar la variable para el total nacional tiene un comportamiento más estable que la variable provincial. Este hecho favorece la proyección. Por otra parte la componente de los diferenciales elimina la influencia transversal sobre la mortalidad de la población debida a circunstancias puntuales climatológicas, epidémicas, etc. Estos factores afectan por igual a las dos poblaciones, tanto nacional como provincial, y son intrínsecos a la coyuntura del año y quedan recogidos en la esperanza de vida para el total nacional. El diferencial recoge el comportamiento particular de cada provincia respecto del total nacional independientemente de los factores comunes que inciden sobre toda la población.

Para especificar los diferenciales de la esperanza de vida provincial se propone el siguiente modelo:

$$re(t, x, h) = a(x, h) + b(x, h) \cdot t + \varepsilon(t, x, h) \quad \text{para } t=1,2,\dots,18 \quad (4.29)$$

Donde $t=1$ es el año 1991, etc. Se hace un cambio de localización para limitar la influencia de la pendiente en el ajuste de este coeficiente. La ordenada en el origen $a(x, h)$ es un parámetro particularmente sensible a la localización de la nube de puntos respecto al eje de ordenadas.

Podemos construir un intervalo de predicción para una observación futura t_0 empezando por calcular la varianza de $re(t_0, x, h) - \hat{r}(t_0, x, h)$:

$$Var[re(t_0, x, h) - \hat{r}(t_0, x, h)] = \sigma^2 + \sigma^2 \left\{ \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{XX}} \right\} = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{XX}} \right\} \quad (4.30)$$

La expresión $\frac{re(t_0, x, h) - \hat{r}(t_0, x, h)}{\sigma \cdot \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{XX}}}}$ sigue una distribución normal tipificada.

A partir de la suma de cuadrados de los residuos de la tabla del Análisis de la Varianza se estima la varianza σ^2 con el estadístico:

$$S^2 = \frac{\|re(t, x, h) - \hat{r}(t, x, h)\|^2}{n-2} \quad (4.31)$$

Sea $t_{1-\alpha/2}$ tal que $P\left[|T| < t_{1-\alpha/2}\right] = 1 - \alpha$ siendo T una variable aleatoria con distribución t de Student con n-2 grados de libertad, sustituyendo σ por S, (ver Graybill (1976)), se tiene que:

$$\left| re(t, x, h) - \hat{a}(x, h) - \hat{b}(x, h) \cdot t \right| \leq \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{XX}}} \cdot S \cdot t_{1-\alpha/2} \quad (4.32)$$

El modelo proporciona el siguiente intervalo de confianza para el diferencial:

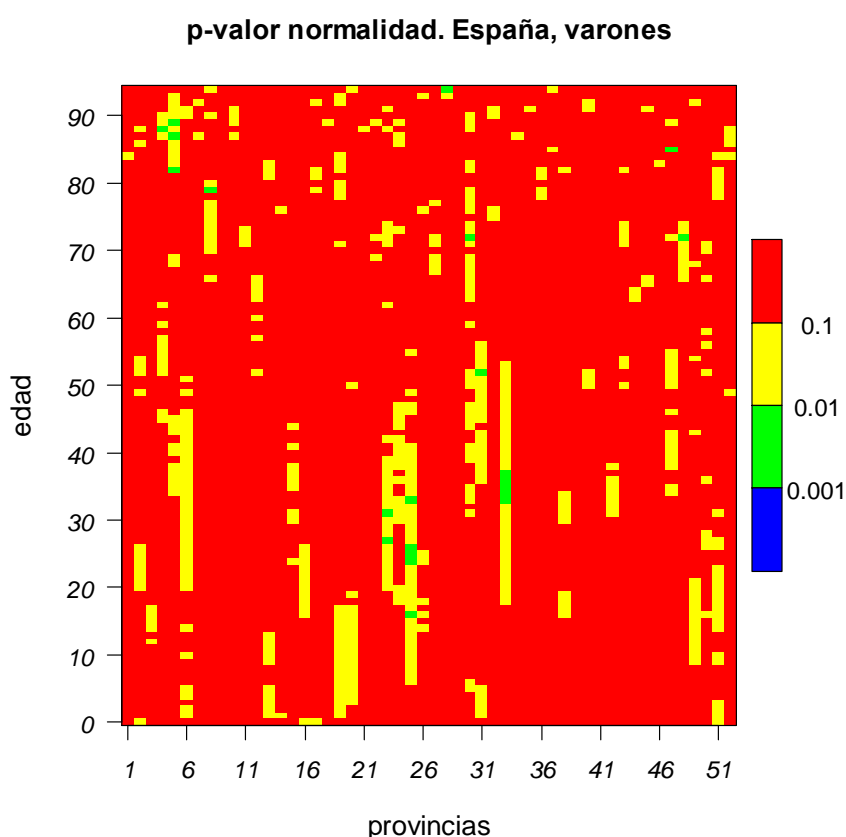
$$re(t_0, x, h) \in \left[\hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{XX}}} \cdot S \cdot t_{1-\alpha/2} \right] \quad (4.33)$$

y de forma inmediata, utilizando las expresiones 4.28 y 4.33, se puede derivar la expresión del intervalo de confianza para la esperanza de vida provincial.

Antes de continuar, es necesario realizar un análisis de residuales para verificar la normalidad y la media nula de los residuos. De la hipótesis de normalidad y media nula dependen las propiedades de los estadísticos e inferencias que se deduzcan de los parámetros.

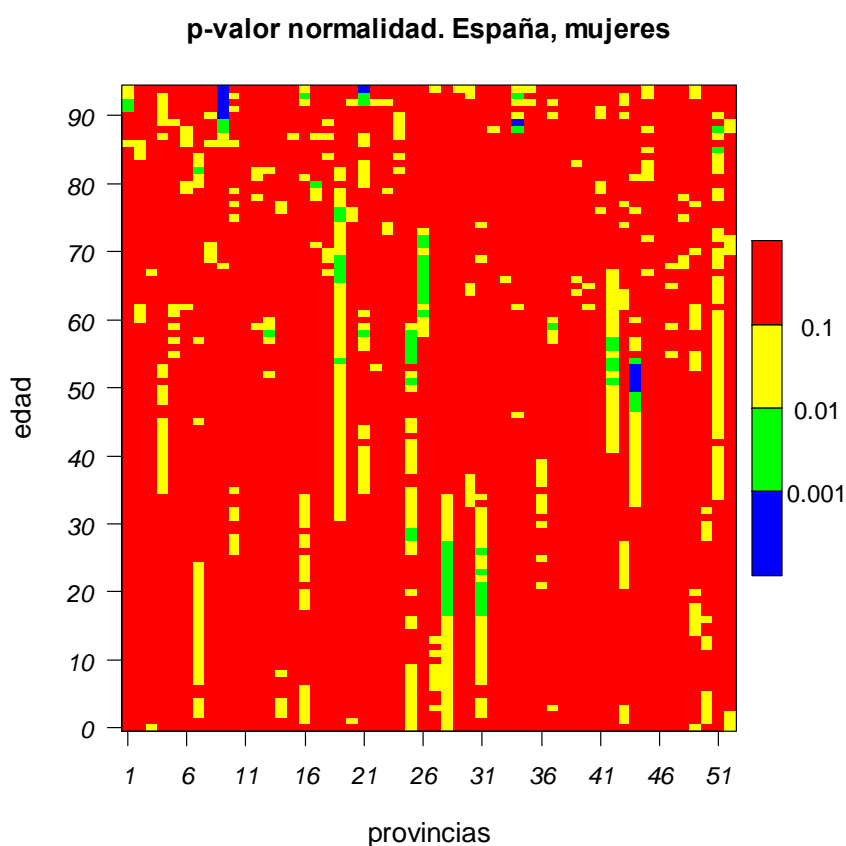
De acuerdo con la Figura 4.1 y la Figura 4.2, que muestra el p-valor para el test de normalidad de los residuos, se acepta la hipótesis nula, tanto para varones como para mujeres.

Figura 4.1. P-valor en el test de normalidad de Shapiro-Wilks para los residuos de los diferenciales de la esperanza de vida provincial. Varones



Fuente: Elaboración propia a partir de datos del INE

Figura 4.2 P-valor en el test de normalidad de Shapiro-Wilks para los residuos de los diferenciales de la esperanza de vida provincial. Mujeres



Fuente: Elaboración propia a partir de datos del INE

En la Figura 4.1 y la Figura 4.2 se observa que la distribución de de los p-valores no depende del sexo ni de la provincia. Las franjas verticales corresponden a provincias y son imputables a la dependencia entre edades que genera la propia construcción de las esperanza de vida por definición. Por otra parte la frecuencia relativa de las 9964 muestras de varones y mujeres que dan por debajo de 0,001 es 0,0013047 la frecuencia por debajo del 0,01 es 0,01053794 y por debajo de 0,1 es 0,11672019. En todos los casos razonablemente próximos. Estos datos corroboran la hipótesis de normalidad del modelo.

Una vez que se acepta la normalidad del modelo es preciso verificar si la media de los residuos es nula. Se realiza un contraste t de Student aplicado a cada una de las muestras de 18 observaciones de cada una de las series para cada provincia, sexo y edad y se confirma que no hay evidencia suficiente para rechazar que la media de los

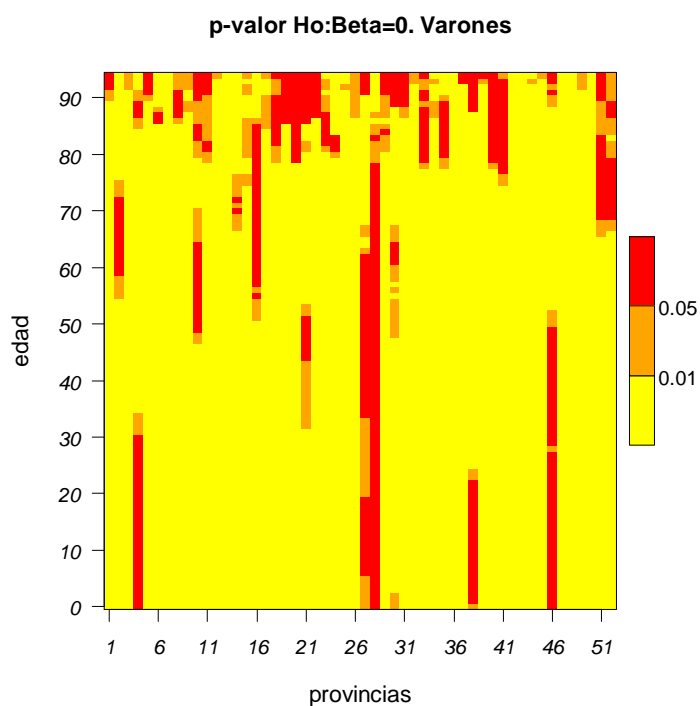
residuos para ninguna de las muestras sea nula. Es más, el p-valor es tan alto que el paquete estadístico SAS da valores de 1 en todos los casos tanto para varones como para mujeres.

El análisis de residuos es satisfactorio en el sentido de que acepta las hipótesis del modelo lineal. Ahora bien, otra cuestión es que tenga sentido ajustar los residuos para que se detecte una tendencia. Es necesario verificar que los parámetros del modelo son significativos conjuntamente. Es decir, que tiene sentido plantear un modelo para explicar la esperanza de vida provincial a partir de la esperanza de vida nacional y una tendencia lineal que sea significativa.

Se trata de analizar los parámetros del modelo mediante un estudio de significación del mismo, no sólo del coeficiente regresor de la variable temporal sino también del término independiente. Es decir se trata de contrastar la hipótesis de que la esperanza de vida provincial se pueda explicar en base a la nacional dentro de un intervalo de confianza dependiente de la provincia.

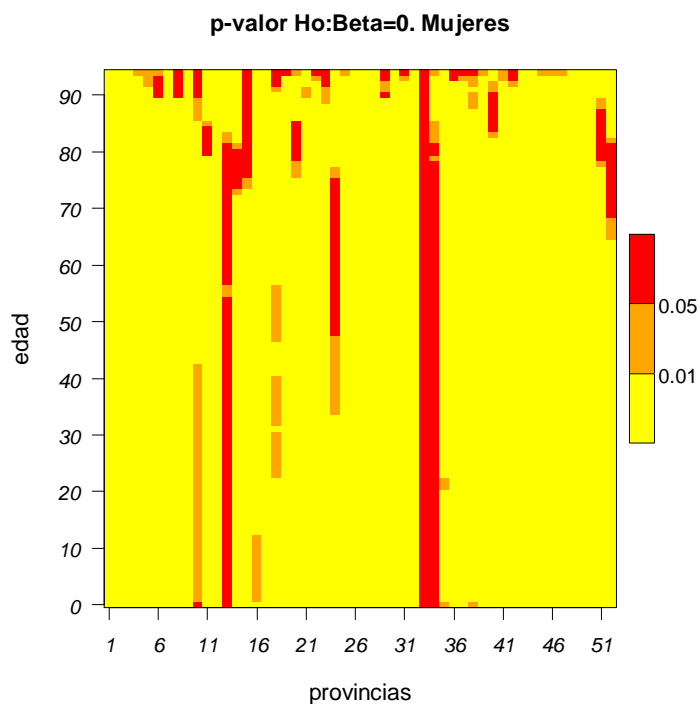
En la Figura 4.3 y en la Figura 4.4 se muestran los resultados del contraste para verificar si los coeficientes son significativos. Como los estadísticos de este contraste no son los que aparecen tradicionalmente en la salida de los ajustes en las regresiones se desarrolló en R un programa “ad hoc”. Las provincias están ordenadas de mayor a menor población para poder detectar una hipotética influencia del tamaño de la población, lo que finalmente se descarta.

Figura 4.3. Superficie con p-valor para el contraste de hipótesis de coeficientes nulos. Varones.



Fuente: Elaboración propia a partir de datos del INE

Figura 4.4. Superficie con p-valor para el contraste de hipótesis de coeficientes nulos. Mujeres.



Fuente: Elaboración propia a partir de datos del INE

El contraste de significación para los coeficientes del parámetro rechaza que sean nulos los dos a la vez de acuerdo con la superficie, tanto para varones como para mujeres, según la Figura 4.3 y la Figura 4.4. Únicamente en algunas franjas verticales correspondientes a provincias y que no coinciden para los dos sexos se acepta la hipótesis nula para un nivel de significación de hasta un mínimo de 0,05. En el caso de los varones se trata de las provincias de Alicante, Castellón, Cantabria, Lugo y Palencia. En las mujeres Asturias, Lérida y Cáceres.

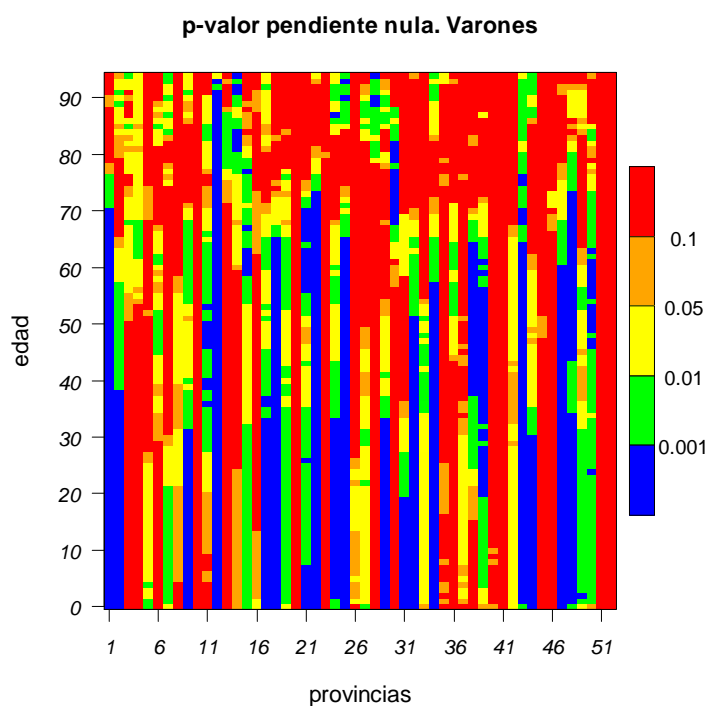
También en la región de las edades más avanzadas, sobre todo en varones, hay mayor propensión a aceptar la hipótesis nula.

Por tanto, de acuerdo con los datos observados, hay fuerte evidencia de que los parámetros son significativos salvo en pocos casos. Es decir, los diferenciales tienen una tendencia lineal significativa y no son ruido blanco y por tanto tiene sentido plantear el modelo que explica la esperanza de vida provincial a partir de la nacional y cierta tendencia lineal.

Aunque, de acuerdo con los datos, alguno de los coeficientes es significativo, es necesario verificar si el diferencial provincial respecto al total nacional se conserva en el tiempo. Para este caso se utiliza el coeficiente de la pendiente o lo que es lo mismo la significación del regresor del modelo.

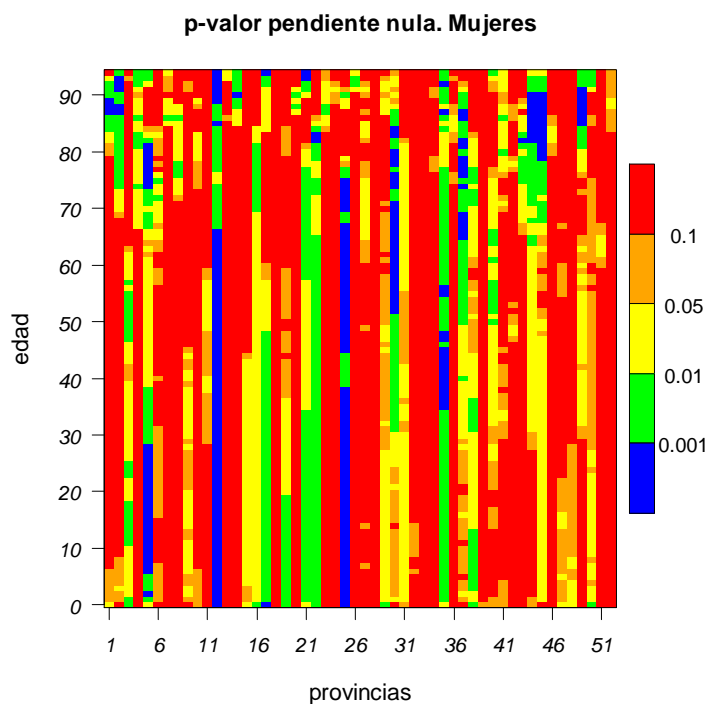
Análogamente, la Figura 4.5 y la Figura 4.6 representan el p-valor del contraste de significación de la variable independiente, es decir el tiempo.

Figura 4.5. Superficie con el p-valor de contraste de hipótesis para la significación del regresor. Varones.



Fuente: Elaboración propia a partir de datos del INE

Figura 4.6. Superficie con el p-valor de contraste de hipótesis para la significación del regresor. Mujeres.



Fuente: Elaboración propia a partir de datos del INE

Así en general, de entre todas las provincias para las que hay evidencia suficiente para rechazar que la hipótesis de que ninguno de los coeficientes del modelo son significativos, es decir, que al menos uno de los coeficientes es significativo se contrasta la hipótesis de que el coeficiente de la variable independiente, en este caso el tiempo, es significativo.

De acuerdo con la Figura 4.5 y la Figura 4.6, para un nivel de significación de 0,05 el test no encuentra evidencia suficiente para rechazar la hipótesis nula, es decir, que la pendiente de la recta es cero, con lo que el número total de provincias para las que la pendiente de la recta es nula es de 26 para varones y 35 para mujeres, incluyendo los casos comentados anteriormente en que ambos parámetros se consideran nulos.

Para el conjunto de provincias en el que se acepta que la pendiente sea nula el diferencial de la esperanza de vida respecto al total nacional se vuelve a estimar ajustando un modelo con término constante. Estas provincias tienen un diferencial respecto al total nacional constante, de lo que se deduce que no hay evidencia de una posible convergencia provincial con el valor nacional.

4.6 Resultados y conclusiones

En este capítulo se propone un innovador modelo para la estimación y proyección de la mortalidad en las poblaciones de pequeño tamaño. Frente a la bibliografía tradicional que proporciona una estimación puntual, este modelo, además, obtiene una estimación por intervalos que permite medir la bondad de la predicción. Siguiendo el principio de Keynes, "*Es mejor acertar aproximadamente que equivocarse exactamente*", el modelo establece un intervalo de confianza para la estimación y la proyección de valores futuros de las variables provinciales como alternativa a los modelos clásicos.

Se consolida la base teórica, en el apartado 4.2, que subyace bajo las series de tiempo definidas por las funciones biométricas de una tabla de mortalidad, describiendo las series de tiempo como la realización de procesos estocásticos.

A continuación, se acomete una revisión de los métodos de proyección que históricamente se han llevado a cabo en España cuya producción fomenta la incesante demanda de información cada vez más desagregada territorialmente. Esta revisión se acomete en el apartado 4.3. Desde los primeros intentos para la estimación del tamaño de la población del total Nacional de España con las curvas logísticas pasando por los primeros ejercicios con desagregación territorial por provincia a partir de las proyecciones calculadas como coeficientes de mejora hasta el modelo actual que se ha impuesto basado en el criterio relacional entre la mortalidad del total nacional y la mortalidad de cada provincia.

Como extensión de la filosofía de los modelos relacionales se propone un modelo para el riesgo de muerte provincial que se describe en el apartado 4.4. Frente a los inconvenientes ya comentados en el capítulo 3, este modelo da respuesta a las especiales características del riesgo de muerte provincial observado y ofrece una alternativa para la estimación en poblaciones de pequeño tamaño. Esta alternativa deriva en un modelo probabilístico, después de un arduo trabajo en la búsqueda de un solución satisfactoria y que finalmente ha dado sus frutos con una estimación por intervalos para el riesgo de muerte en presencia de riesgo nulo.

Finalmente, se concluye con un modelo para la esperanza de vida provincial, en el apartado 4.5, con mejores resultados que para el riesgo de muerte. Como el modelo para el riesgo de muerte, este modelo está basado en la filosofía de los modelos relacionales.

El modelo se reduce eventualmente a un modelo lineal que responde bien a las expectativas. El análisis de residuales confirma que el ajuste es significativamente bueno. Además, el contraste de significación de los parámetros indica que al menos uno de los parámetros es significativo salvo algún caso excepcional.

Una vez especificado un modelo que parece satisfactorio como es el modelo para la esperanza de vida, queda validarlo con datos de los años 2009, 2010 y 2011 ya conocidos. Este es el propósito del siguiente capítulo.

Capítulo 5

Validación del Intervalo de confianza del modelo lineal para la esperanza de vida provincial

- 5.1 Introducción
- 5.2 Esperanza de vida al nacimiento observada en 2009, 2010 y 2011
- 5.3 Éxito del intervalo de confianza con la esperanza de vida observada
- 5.4 Éxito del intervalo de confianza con las proyecciones a corto plazo
- 5.5 Resultados y conclusiones

5.1 Introducción

En el capítulo 4 se ha especificado un modelo lineal para los diferenciales provinciales de la esperanza de vida, $re(t,x,h)$, definido como la diferencia entre la esperanza de vida provincial y la esperanza de vida nacional para cada edad y año. El modelo se aplica al periodo de tiempo que va desde el año 1991 hasta 2008. De acuerdo con los resultados de los contrastes de hipótesis en el análisis de residuales del modelo no hay evidencia suficiente para rechazar que se cumplen todas las hipótesis del modelo lineal:

1. Aditividad del modelo: el diferencial $e(t,x,h)$ se desagrega en una parte determinista lineal que depende del tiempo y una parte aleatoria, $re(t,x,h)$.
2. La parte aleatoria sigue una distribución normal de media nula y varianza σ^2 .

En este capítulo se va a estudiar el éxito del intervalo de confianza proporcionado por el modelo propuesto en el apartado 5 del capítulo 4. Se va a validar el modelo frente a los datos observados y se van a comparar sus proyecciones con los datos de las proyecciones a corto plazo elaboradas por el INE.

En particular se seguirá un estudio atendiendo a las distintas variables implicadas como son el sexo, la edad y la provincia.

Este capítulo se desarrolla en 5 apartados.

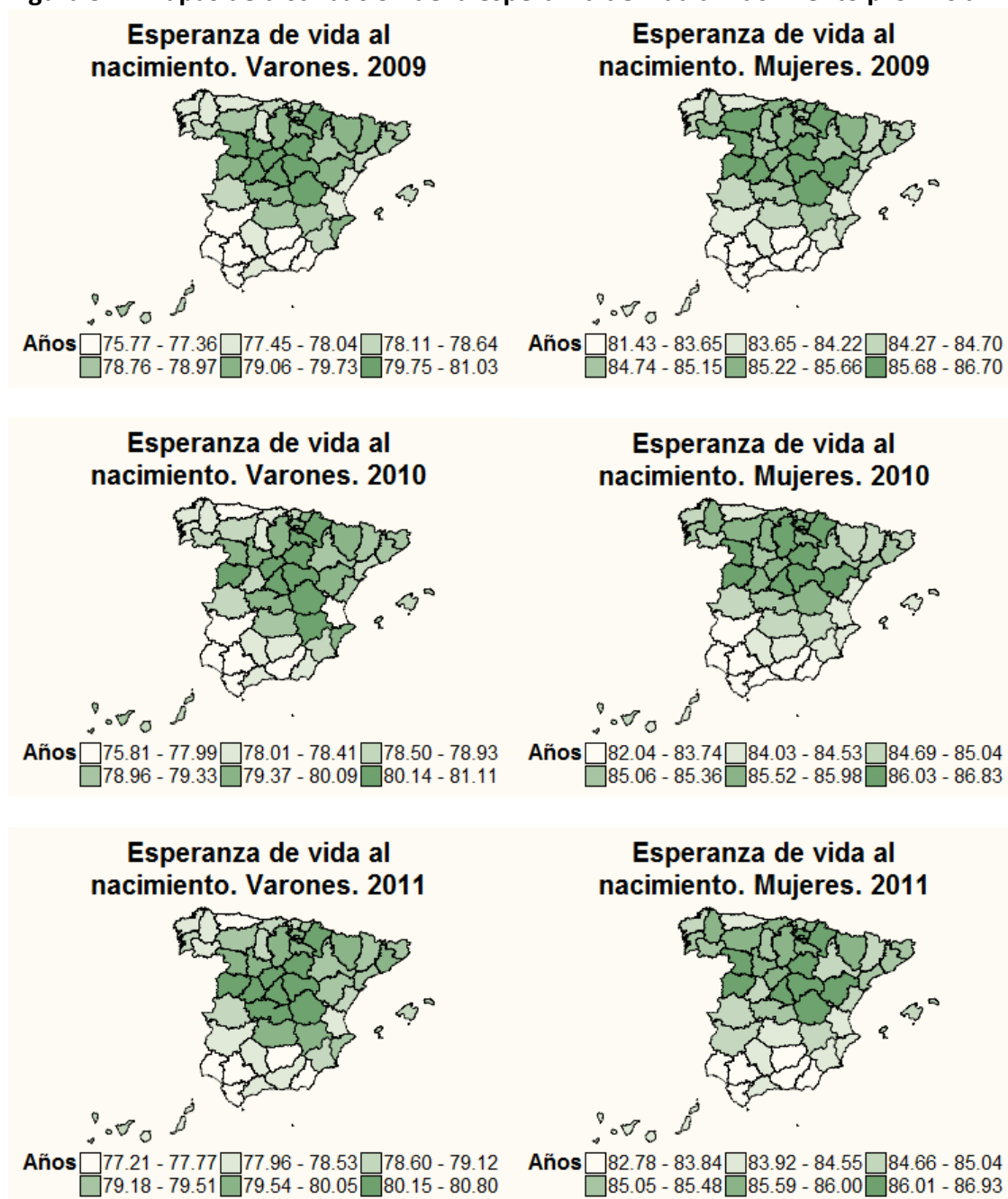
- En el segundo apartado, se presentan la esperanza de vida al nacimiento observada durante los años 2009, 2010 y 2011. Estos datos son continuación de la serie de datos observados a partir de los cuales se ha ajustado el modelo del apartado 4.5.
- El tercero consiste en un estudio del éxito del intervalo de confianza predicho a partir de los datos observados y descritos en el apartado anterior.
- El cuarto se ocupa de un estudio comparativo entre el modelo propuesto en el apartado 4.5 y la esperanza de vida al nacimiento de las proyecciones a corto plazo elaboradas por el INE

- Se termina con un quinto apartado donde se recogen los resultados del capítulo y se exponen las conclusiones a las que conducen, valorando positivamente el alto grado de cumplimiento de las metas buscadas con el trabajo realizado.

5.2 Esperanza de vida al nacimiento observada en 2009, 2010 y 2011

En la Figura 5.1 se muestran los mapas de la distribución de la esperanza de vida al nacimiento por sexo en las provincias españolas durante los años 2009, 2010 y 2011. En ellos se aprecia el mismo patrón observado en el capítulo 3 para la serie desde 1991 hasta 2008. Las provincias del interior tienen mayor esperanza de vida que las costeras. Al mismo tiempo la esperanza de vida al nacimiento va aumentando de sur a norte y de oeste a este, y se confirma que la distribución relativa geográfica es independiente del sexo.

Figura 5.1. Mapas de distribución de la esperanza de vida al nacimiento provincial.



Fuente: Elaboración propia a partir de datos del INE

En la Tabla 5.1 se muestra el detalle de los valores representados en los mapas de la esperanza de vida al nacimiento para los varones de la Figura 5.1.

**Tabla 5.1. Esperanza de vida al nacimiento nacional y provincial.
Esperanza de vida al nacimiento observada en 2009, 2010 y 2011**

| Provincias | Varones | | | Mujeres | | |
|------------------------|---------|-------|-------|---------|-------|-------|
| | 2009 | 2010 | 2011 | 2009 | 2010 | 2011 |
| Total Nacional | 78,55 | 78,94 | 79,16 | 84,56 | 84,91 | 84,97 |
| Araba/Alava | 79,21 | 80,02 | 79,69 | 85,15 | 86,03 | 86,01 |
| Albacete | 78,88 | 80,16 | 79,81 | 84,74 | 84,91 | 84,88 |
| Alicante/Alacant | 79,10 | 79,37 | 79,46 | 84,43 | 84,44 | 84,53 |
| Almería | 77,20 | 78,02 | 77,70 | 83,09 | 83,50 | 83,58 |
| Ávila | 80,21 | 78,85 | 80,33 | 85,74 | 85,67 | 84,86 |
| Badajoz | 77,11 | 77,62 | 77,97 | 84,07 | 83,66 | 84,74 |
| Balears, Illes | 78,41 | 78,80 | 79,10 | 83,95 | 84,45 | 84,79 |
| Barcelona | 78,97 | 79,25 | 79,54 | 84,81 | 85,22 | 85,36 |
| Burgos | 79,18 | 80,09 | 80,05 | 85,58 | 86,24 | 85,67 |
| Cáceres | 78,60 | 78,52 | 78,84 | 84,64 | 84,90 | 84,89 |
| Cádiz | 76,81 | 77,22 | 77,33 | 82,86 | 83,27 | 83,20 |
| Castellón/Castelló | 77,50 | 79,07 | 78,89 | 84,27 | 84,46 | 84,66 |
| Ciudad Real | 78,76 | 79,31 | 79,55 | 84,28 | 85,00 | 84,22 |
| Córdoba | 77,68 | 78,12 | 78,41 | 84,06 | 84,39 | 84,23 |
| Coruña, A | 77,67 | 78,24 | 78,43 | 84,22 | 84,80 | 85,04 |
| Cuenca | 80,55 | 80,31 | 80,15 | 85,82 | 85,53 | 86,10 |
| Girona | 78,79 | 79,22 | 79,43 | 84,68 | 85,00 | 85,39 |
| Granada | 77,27 | 77,82 | 78,53 | 83,51 | 83,59 | 83,48 |
| Guadalajara | 81,03 | 81,03 | 80,75 | 86,30 | 86,20 | 86,28 |
| Gipuzkoa | 78,84 | 79,57 | 79,28 | 85,32 | 85,60 | 85,84 |
| Huelva | 77,36 | 77,99 | 77,77 | 83,15 | 83,25 | 83,45 |
| Huesca | 79,06 | 79,68 | 79,80 | 85,25 | 84,69 | 85,59 |
| Jaén | 76,91 | 78,01 | 77,68 | 83,65 | 84,34 | 83,84 |
| León | 78,92 | 78,82 | 79,51 | 85,88 | 85,98 | 85,86 |
| Lleida | 79,12 | 78,96 | 79,32 | 84,45 | 84,88 | 85,00 |
| Rioja, La | 79,55 | 80,14 | 79,84 | 85,52 | 85,70 | 85,87 |
| Lugo | 77,46 | 78,13 | 78,48 | 84,69 | 85,52 | 85,59 |
| Madrid | 79,92 | 80,46 | 80,74 | 85,66 | 86,04 | 86,15 |
| Málaga | 77,71 | 78,02 | 78,45 | 83,28 | 83,74 | 83,92 |
| Murcia | 78,18 | 78,93 | 78,92 | 83,65 | 84,28 | 84,44 |
| Navarra | 79,75 | 80,76 | 80,76 | 86,03 | 86,53 | 86,06 |
| Ourense | 78,64 | 78,81 | 77,96 | 85,43 | 85,04 | 85,31 |
| Asturias | 77,45 | 77,29 | 77,50 | 84,13 | 84,53 | 84,55 |
| Palencia | 77,89 | 78,22 | 79,12 | 84,95 | 85,66 | 85,48 |
| Palmas, Las | 78,11 | 78,50 | 78,50 | 84,12 | 84,40 | 84,22 |
| Pontevedra | 78,04 | 78,41 | 78,79 | 84,70 | 85,27 | 85,21 |
| Salamanca | 79,69 | 80,41 | 80,56 | 85,78 | 86,47 | 86,72 |
| Santa Cruz de Tenerife | 78,96 | 78,97 | 78,60 | 83,72 | 84,83 | 84,55 |
| Cantabria | 78,40 | 78,33 | 78,87 | 85,31 | 85,54 | 85,61 |
| Segovia | 80,72 | 80,65 | 80,56 | 85,39 | 85,94 | 86,00 |
| Sevilla | 77,06 | 77,26 | 77,73 | 83,21 | 83,44 | 83,30 |
| Soria | 79,99 | 81,11 | 80,69 | 86,70 | 86,23 | 86,93 |
| Tarragona | 78,35 | 78,83 | 79,33 | 84,63 | 85,06 | 85,05 |
| Teruel | 79,22 | 79,42 | 79,43 | 85,68 | 86,07 | 86,32 |
| Toledo | 79,73 | 80,09 | 80,80 | 84,94 | 85,27 | 85,23 |
| Valencia/València | 77,86 | 77,95 | 78,33 | 83,86 | 84,03 | 84,21 |
| Valladolid | 80,11 | 79,42 | 79,84 | 85,09 | 85,36 | 85,85 |
| Bizkaia | 78,16 | 78,62 | 78,90 | 85,22 | 85,36 | 85,20 |
| Zamora | 79,83 | 80,02 | 79,99 | 85,94 | 86,83 | 86,50 |
| Zaragoza | 78,81 | 79,06 | 79,18 | 84,89 | 85,07 | 85,01 |
| Ceuta | 75,77 | 75,81 | 77,21 | 81,43 | 82,04 | 82,78 |
| Melilla | 78,42 | 79,33 | 77,44 | 84,93 | 85,12 | 83,58 |

Fuente: Elaboración propia a partir de datos del INE

Se puede observar que los valores para el total nacional presentan un crecimiento monótono. Por el contrario las provincias presentan en general una tendencia creciente, aunque en algunos casos fluctúa de año en año, sobre todo en las provincias de pequeño tamaño poblacional.

5.3 Éxito del intervalo de confianza con la esperanza de vida observada

Para cada año y edad, la expresión 4.33 permite determinar un intervalo de confianza para la esperanza de vida en cada provincia. Dado un nivel de confianza del 95% se tienen 29.892¹ intervalos de confianza que testar. Se define el éxito como que los diferenciales de la esperanza de vida observada, definidos por la fórmula 4.27, caigan dentro del intervalo de confianza predicho por la fórmula 4.33 para los 3 años de prueba, las 52 provincias, los 2 sexos y las edades. El recuento total de éxitos es 27.687, lo que supone un 92,62%.

De acuerdo con la Tabla 5.2, el porcentaje de acierto por año está por encima del 90% siendo de 92,11% en 2009, 90,26% en 2010 y 95,49% en 2011. Sin una tendencia clara, el mejor resultado se obtiene en el año 2011.

**Tabla 5.2. Porcentaje de éxito por sexo y año para el intervalo provincial
Porcentaje de éxito por sexo y año al 95%**

| | 2009 | 2010 | 2011 | Total |
|---------|--------|--------|--------|--------|
| Varones | 88,96% | 90,79% | 94,78% | 91,51% |
| Mujeres | 95,26% | 89,74% | 96,21% | 93,74% |
| Total | 92,11% | 90,26% | 95,49% | 92,62% |

Fuente: Elaboración propia a partir de datos del INE

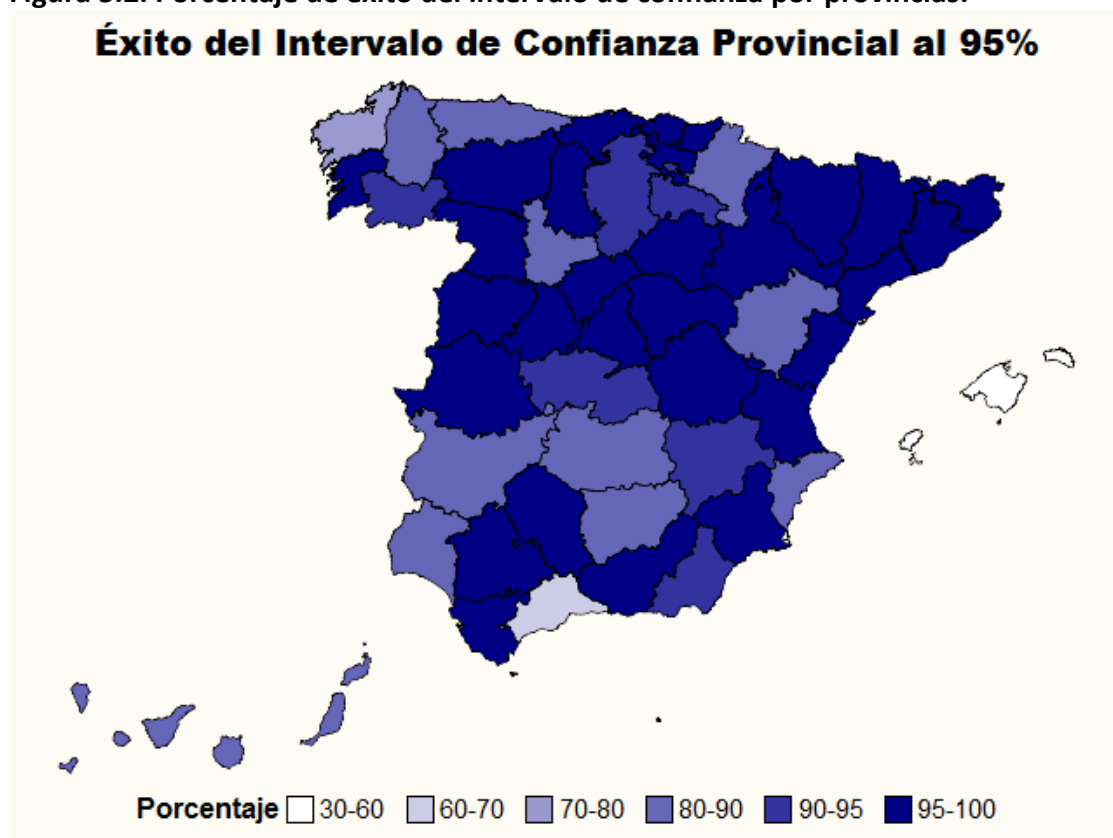
Por sexos el porcentaje es superior para mujeres con un 93,74% frente a los varones con un 91,51%, en sintonía con la mayor estabilidad de las series observadas, en el caso de las mujeres.

¹ 3 años x 2 sexos x 96 edades x 50 provincias + 3 años x 2 sexos x 91 edades x 2 Ciudades Autónomas.

Son de particular interés los datos obtenidos hasta el año 2010, porque se trata del año actual desde la perspectiva de la simulación del ejercicio. Las *Estimaciones de Población Actual* se construyen a partir de estos datos.

La distribución geográfica del porcentaje de éxito es aleatoria como se puede ver en la Figura 5.2. No es posible distinguir regiones aunque sí, en general, un porcentaje alto a excepción de unos pocos casos.

Figura 5.2. Porcentaje de éxito del intervalo de confianza por provincias.



Fuente: Elaboración propia a partir de datos del INE

En la Tabla 5.3 se comprueba que el porcentaje de éxito para el intervalo de confianza está, salvo para dos provincias, Baleares con un 32,12% y Málaga con un 63,19%, ampliamente por encima del 70%. La proporción es independiente del tamaño de la provincia pues estas dos provincias son grandes comparativamente. Por otra parte, una población de pequeño tamaño como Melilla tiene un porcentaje alto de éxito. Hay 30 provincias que superan el 95% y la mediana del porcentaje de éxito es del 97,02%.

En la Tabla 5.3 se presenta también el detalle de la distribución del porcentaje de éxito por sexo. En 36 casos el porcentaje de éxito alcanza el 100%, para varones o para mujeres.

Tabla 5.3. Porcentaje de éxito del intervalo de confianza por provincia.

Porcentaje de éxito para el intervalo de confianza por provincias al 95%

| Provincias | Varones | Mujeres | Total | Provincias | Varones | Mujeres | Total |
|--------------------|---------|---------|---------|------------------------|---------------|---------------|---------------|
| Araba/Alava | 99,31% | 99,31% | 99,31% | Navarra | 64,24% | 98,96% | 81,60% |
| Albacete | 80,56% | 100,00% | 90,28% | Ourense | 100,00% | 81,25% | 90,63% |
| Alicante/Alacant | 71,88% | 97,22% | 84,55% | Asturias | 76,04% | 100,00% | 88,02% |
| Almería | 85,07% | 100,00% | 92,53% | Palencia | 99,31% | 98,26% | 98,78% |
| Ávila | 99,65% | 100,00% | 99,83% | Palmas, Las | 77,08% | 97,22% | 87,15% |
| Badajoz | 74,31% | 94,79% | 84,55% | Pontevedra | 100,00% | 97,92% | 98,96% |
| Balears, Illes | 31,25% | 32,99% | 32,12% | Salamanca | 97,22% | 97,92% | 97,57% |
| Barcelona | 97,92% | 99,65% | 98,78% | Santa Cruz de Tenerife | 78,82% | 88,89% | 83,85% |
| Burgos | 100,00% | 87,15% | 93,58% | Cantabria | 95,49% | 100,00% | 97,74% |
| Cáceres | 97,92% | 100,00% | 98,96% | Segovia | 97,22% | 97,92% | 97,57% |
| Cádiz | 100,00% | 100,00% | 100,00% | Sevilla | 99,65% | 93,40% | 96,53% |
| Castellón/Castelló | 99,31% | 100,00% | 99,65% | Soria | 98,61% | 100,00% | 99,31% |
| Ciudad Real | 100,00% | 73,61% | 86,81% | Tarragona | 100,00% | 100,00% | 100,00% |
| Córdoba | 100,00% | 98,26% | 99,13% | Teruel | 80,90% | 96,18% | 88,54% |
| Coruña, A | 100,00% | 56,60% | 78,30% | Toledo | 84,72% | 100,00% | 92,36% |
| Cuenca | 99,65% | 97,57% | 98,61% | Valencia/València | 99,65% | 100,00% | 99,83% |
| Girona | 99,31% | 100,00% | 99,65% | Valladolid | 73,61% | 99,65% | 86,63% |
| Granada | 100,00% | 97,57% | 98,78% | Bizkaia | 93,75% | 97,92% | 95,83% |
| Guadalajara | 100,00% | 99,31% | 99,65% | Zamora | 97,22% | 100,00% | 98,61% |
| Gipuzkoa | 100,00% | 100,00% | 100,00% | Zaragoza | 93,06% | 97,22% | 95,14% |
| Huelva | 76,39% | 100,00% | 88,19% | Ceuta | 88,64% | 94,79% | 91,72% |
| Huesca | 100,00% | 100,00% | 100,00% | Melilla | 100,00% | 94,79% | 97,40% |
| Jaén | 99,31% | 71,18% | 85,24% | Total | 91,51% | 93,74% | 92,62% |
| León | 100,00% | 100,00% | 100,00% | | | | |
| Lleida | 97,22% | 100,00% | 98,61% | | | | |
| Rioja, La | 99,31% | 82,99% | 91,15% | | | | |
| Lugo | 73,96% | 100,00% | 86,98% | | | | |
| Madrid | 100,00% | 100,00% | 100,00% | | | | |
| Málaga | 81,94% | 44,44% | 63,19% | | | | |
| Murcia | 99,31% | 99,65% | 99,48% | | | | |

Fuente: Elaboración propia a partir de datos del INE

Para las provincias de Cádiz, Guipúzcoa, León, Madrid y Tarragona el éxito es total. Por otra parte, no se aprecia ningún patrón geográfico ni correlación con el tamaño de la población. En particular, Melilla o Soria alcanzan un significativo porcentaje del 97,4% o 99,3% de éxito, respectivamente.

La distribución por edades se presenta en la Tabla 5.4. Se puede apreciar que el porcentaje de acierto del intervalo es superior al 90% en todas las edades con un mínimo en las primeras edades. El porcentaje de éxito va subiendo hasta alcanzar su máximo para las edades en torno a los 86 años, en alta correlación con el número de defunciones teórico $d(t, x)$ como se puede ver en la Figura 3.3 del capítulo 3.

Tabla 5.4. Porcentaje de éxito por edad y sexo para el intervalo provincial al 95%
Porcentaje de éxito por edad para un intervalo del 95%

| Edad | Varones | Mujeres | Total | Edad | Varones | Mujeres | Total |
|------|---------|---------|--------|------|---------|---------|--------|
| 0 | 85,90% | 95,51% | 90,71% | 48 | 91,03% | 94,23% | 92,63% |
| 1 | 87,82% | 92,31% | 90,06% | 49 | 91,67% | 93,59% | 92,63% |
| 2 | 87,18% | 92,31% | 89,74% | 50 | 91,67% | 92,31% | 91,99% |
| 3 | 88,46% | 92,95% | 90,71% | 51 | 91,67% | 92,95% | 92,31% |
| 4 | 88,46% | 92,95% | 90,71% | 52 | 89,74% | 92,95% | 91,35% |
| 5 | 89,10% | 92,31% | 90,71% | 53 | 91,67% | 94,23% | 92,95% |
| 6 | 89,10% | 92,31% | 90,71% | 54 | 92,31% | 94,23% | 93,27% |
| 7 | 89,10% | 92,95% | 91,03% | 55 | 91,67% | 94,23% | 92,95% |
| 8 | 87,82% | 92,95% | 90,38% | 56 | 90,38% | 93,59% | 91,99% |
| 9 | 88,46% | 92,95% | 90,71% | 57 | 91,03% | 93,59% | 92,31% |
| 10 | 88,46% | 92,95% | 90,71% | 58 | 90,38% | 93,59% | 91,99% |
| 11 | 87,82% | 92,95% | 90,38% | 59 | 89,74% | 93,59% | 91,67% |
| 12 | 88,46% | 92,95% | 90,71% | 60 | 92,31% | 94,23% | 93,27% |
| 13 | 89,10% | 92,95% | 91,03% | 61 | 94,23% | 94,23% | 94,23% |
| 14 | 87,82% | 92,31% | 90,06% | 62 | 94,23% | 94,23% | 94,23% |
| 15 | 87,18% | 92,31% | 89,74% | 63 | 95,51% | 94,87% | 95,19% |
| 16 | 87,18% | 92,31% | 89,74% | 64 | 94,87% | 94,87% | 94,87% |
| 17 | 87,18% | 92,95% | 90,06% | 65 | 94,23% | 93,59% | 93,91% |
| 18 | 88,46% | 92,95% | 90,71% | 66 | 94,87% | 94,23% | 94,55% |
| 19 | 87,82% | 92,95% | 90,38% | 67 | 94,87% | 94,23% | 94,55% |
| 20 | 89,10% | 92,31% | 90,71% | 68 | 96,79% | 92,31% | 94,55% |
| 21 | 89,74% | 92,95% | 91,35% | 69 | 94,23% | 91,67% | 92,95% |
| 22 | 87,82% | 92,31% | 90,06% | 70 | 94,87% | 93,59% | 94,23% |
| 23 | 87,82% | 93,59% | 90,71% | 71 | 97,44% | 90,38% | 93,91% |
| 24 | 87,18% | 92,95% | 90,06% | 72 | 94,87% | 90,38% | 92,63% |
| 25 | 88,46% | 93,59% | 91,03% | 73 | 96,15% | 90,38% | 93,27% |
| 26 | 87,82% | 92,95% | 90,38% | 74 | 94,87% | 92,95% | 93,91% |
| 27 | 88,46% | 93,59% | 91,03% | 75 | 94,23% | 91,67% | 92,95% |
| 28 | 89,74% | 93,59% | 91,67% | 76 | 95,51% | 92,95% | 94,23% |
| 29 | 90,38% | 93,59% | 91,99% | 77 | 94,23% | 92,31% | 93,27% |
| 30 | 90,38% | 93,59% | 91,99% | 78 | 93,59% | 94,23% | 93,91% |
| 31 | 90,38% | 93,59% | 91,99% | 79 | 94,87% | 95,51% | 95,19% |
| 32 | 89,10% | 93,59% | 91,35% | 80 | 91,67% | 96,15% | 93,91% |
| 33 | 89,74% | 94,23% | 91,99% | 81 | 92,95% | 94,23% | 93,59% |
| 34 | 89,74% | 93,59% | 91,67% | 82 | 92,95% | 94,87% | 93,91% |
| 35 | 89,10% | 94,23% | 91,67% | 83 | 95,51% | 96,15% | 95,83% |
| 36 | 89,10% | 94,23% | 91,67% | 84 | 95,51% | 96,15% | 95,83% |
| 37 | 89,10% | 93,59% | 91,35% | 85 | 97,44% | 94,87% | 96,15% |
| 38 | 90,38% | 92,95% | 91,67% | 86 | 98,08% | 95,51% | 96,79% |
| 39 | 89,74% | 92,95% | 91,35% | 87 | 97,44% | 94,87% | 96,15% |
| 40 | 90,38% | 92,95% | 91,67% | 88 | 96,15% | 94,87% | 95,51% |
| 41 | 89,74% | 93,59% | 91,67% | 89 | 96,79% | 96,15% | 96,47% |
| 42 | 89,10% | 94,23% | 91,67% | 90 | 96,79% | 96,15% | 96,47% |
| 43 | 89,10% | 94,87% | 91,99% | 91 | 92,95% | 92,31% | 92,63% |
| 44 | 87,82% | 94,23% | 91,03% | 92 | 92,31% | 93,59% | 92,95% |
| 45 | 88,46% | 93,59% | 91,03% | 93 | 92,95% | 96,15% | 94,55% |
| 46 | 89,74% | 94,23% | 91,99% | 94 | 94,23% | 94,87% | 94,55% |
| 47 | 91,67% | 91,67% | 91,67% | 95 | 93,59% | 93,59% | 93,59% |

Fuente: Elaboración propia a partir de datos del INE

A diferencia de la distribución por provincia en la distribución por edad, no hay datos atípicos. El mínimo por edades para el total de la población se alcanza a los 15 años con un porcentaje de éxito del 89,74%

5.4 Éxito del intervalo de confianza con las proyecciones a corto plazo

Las Proyecciones de Población a Corto Plazo que elabora el INE anualmente requieren del cálculo de las tasas perspectivas de paso. Estos parámetros se obtienen a partir de la proyección de las tablas de mortalidad tanto nacionales como provinciales de forma análoga a las estimaciones de población actual. El lector interesado en el detalle de las proyecciones puede consultar los manuales metodológicos del INE (2009, 2010, 2011, 2012c, 2013b).

En nuestro ejercicio de validación del modelo propuesto en el apartado 4.5 se utilizarán las esperanzas de vida estimadas para las proyecciones de población a corto plazo elaboradas por el INE. Se aprovecharán los datos de las PPCP (2010-2020). Los datos se han extraído de la página web del INE². En diciembre de 2013 ya estaban publicadas las proyecciones 2013-2023 que incorporan la proyección de las tablas de mortalidad del periodo 2012-2022. Siempre un año menos porque con la tabla de mortalidad del año 2022 se obtiene la población a 1 de enero de 2023. Pero para validar nuestro ejercicio no sirven porque se proyectarán datos observados dos años antes, en nuestro caso el año 2011.

Por tanto, para nuestro ejercicio de validación necesitamos las proyecciones a corto plazo (2010-2020) que incorpora la serie proyectada de Tablas de mortalidad (2009-2019) calculada a partir de datos observados hasta el año 2008 como en nuestro modelo del apartado 4.5.

De acuerdo con la fórmula 4.28, la predicción de la esperanza de vida para cada provincia se obtiene sumando al valor predicho para el total nacional, el

² <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fp269&file=inebase&L=0>.

intervalo de confianza para los diferenciales de la esperanza de vida al nacimiento para cada provincia definido por la fórmula 4.33, según la expresión 5.1.

$$e(t_0, x, h) \in \left[e(t_0, x) + \hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{S_{xx}}} \cdot S \cdot t_{1-\alpha/2} \right] \quad (5.1)$$

En la Tabla 5.5 y la Tabla 5.6 se presentan los extremos del intervalo de confianza de la esperanza de vida proyectada desde 2009 a 2019 para cada provincia:

Tabla 5.5. Extremos del intervalo de confianza provincial al 95% proyectado en el corto plazo para la esperanza de vida al nacimiento. Varones. Periodo 2009 al 2014. Extremos predichos para el intervalo provincial al 95% de la esperanza de vida al nacimiento

| Provincias | 2009 | | 2010 | | 2011 | | 2012 | | 2013 | | 2014 | |
|------------------------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
| | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. |
| Araba/Alava | 78,27 | 80,43 | 78,47 | 80,67 | 78,67 | 80,91 | 78,87 | 81,14 | 79,06 | 81,38 | 79,25 | 81,62 |
| Albacete | 78,57 | 80,03 | 78,74 | 80,23 | 78,91 | 80,43 | 79,08 | 80,62 | 79,24 | 80,81 | 79,40 | 81,00 |
| Alicante/Alacant | 77,29 | 78,52 | 77,46 | 78,72 | 77,64 | 78,92 | 77,81 | 79,11 | 77,98 | 79,31 | 78,14 | 79,50 |
| Almería | 75,55 | 76,95 | 75,65 | 77,07 | 75,75 | 77,19 | 75,84 | 77,31 | 75,93 | 77,43 | 76,02 | 77,55 |
| Ávila | 77,58 | 79,83 | 77,66 | 79,94 | 77,73 | 80,05 | 77,80 | 80,16 | 77,86 | 80,27 | 77,92 | 80,38 |
| Badajoz | 76,82 | 77,94 | 77,01 | 78,14 | 77,19 | 78,34 | 77,36 | 78,54 | 77,54 | 78,73 | 77,70 | 78,92 |
| Balears, Illes | 78,08 | 79,61 | 78,37 | 79,93 | 78,66 | 80,24 | 78,95 | 80,56 | 79,23 | 80,87 | 79,51 | 81,18 |
| Barcelona | 78,40 | 78,92 | 78,64 | 79,17 | 78,88 | 79,41 | 79,11 | 79,66 | 79,34 | 79,90 | 79,56 | 80,13 |
| Burgos | 78,25 | 80,22 | 78,43 | 80,44 | 78,61 | 80,65 | 78,78 | 80,86 | 78,95 | 81,07 | 79,12 | 81,28 |
| Cáceres | 77,41 | 78,47 | 77,53 | 78,62 | 77,66 | 78,76 | 77,78 | 78,90 | 77,90 | 79,04 | 78,01 | 79,18 |
| Cádiz | 75,60 | 77,70 | 75,82 | 77,96 | 76,03 | 78,21 | 76,25 | 78,46 | 76,46 | 78,72 | 76,66 | 78,97 |
| Castellón/Castelló | 76,86 | 78,60 | 77,00 | 78,77 | 77,14 | 78,94 | 77,28 | 79,11 | 77,41 | 79,28 | 77,53 | 79,44 |
| Ciudad Real | 77,38 | 79,20 | 77,55 | 79,41 | 77,72 | 79,61 | 77,88 | 79,81 | 78,04 | 80,01 | 78,20 | 80,20 |
| Córdoba | 76,82 | 78,18 | 76,98 | 78,37 | 77,14 | 78,55 | 77,29 | 78,72 | 77,44 | 78,90 | 77,58 | 79,08 |
| Coruña, A | 76,98 | 78,46 | 77,20 | 78,69 | 77,40 | 78,93 | 77,61 | 79,16 | 77,81 | 79,39 | 78,00 | 79,62 |
| Cuenca | 78,81 | 80,30 | 78,90 | 80,42 | 79,00 | 80,54 | 79,08 | 80,65 | 79,17 | 80,77 | 79,25 | 80,88 |
| Girona | 78,24 | 79,15 | 78,44 | 79,36 | 78,63 | 79,57 | 78,83 | 79,78 | 79,01 | 79,99 | 79,20 | 80,19 |
| Granada | 76,57 | 77,66 | 76,72 | 77,83 | 76,86 | 77,99 | 77,01 | 78,16 | 77,15 | 78,32 | 77,28 | 78,48 |
| Guadalajara | 78,83 | 81,96 | 78,94 | 82,12 | 79,05 | 82,28 | 79,15 | 82,44 | 79,25 | 82,60 | 79,34 | 82,76 |
| Gipuzkoa | 78,02 | 79,40 | 78,27 | 79,67 | 78,52 | 79,95 | 78,77 | 80,22 | 79,01 | 80,48 | 79,24 | 80,75 |
| Huelva | 76,05 | 76,94 | 76,21 | 77,11 | 76,36 | 77,28 | 76,50 | 77,44 | 76,65 | 77,60 | 76,79 | 77,76 |
| Huesca | 77,41 | 80,50 | 77,48 | 80,62 | 77,55 | 80,75 | 77,62 | 80,87 | 77,68 | 80,99 | 77,73 | 81,11 |
| Jaén | 76,38 | 77,90 | 76,49 | 78,03 | 76,59 | 78,16 | 76,69 | 78,29 | 76,79 | 78,42 | 76,88 | 78,54 |
| León | 77,42 | 79,07 | 77,53 | 79,21 | 77,64 | 79,35 | 77,75 | 79,49 | 77,85 | 79,63 | 77,95 | 79,76 |
| Lleida | 77,60 | 79,14 | 77,76 | 79,33 | 77,92 | 79,52 | 78,08 | 79,71 | 78,23 | 79,89 | 78,38 | 80,08 |
| Rioja, La | 78,17 | 79,49 | 78,35 | 79,69 | 78,52 | 79,88 | 78,68 | 80,07 | 78,84 | 80,26 | 79,00 | 80,45 |
| Lugo | 77,21 | 78,84 | 77,41 | 79,07 | 77,61 | 79,29 | 77,80 | 79,52 | 77,99 | 79,74 | 78,18 | 79,97 |
| Madrid | 79,15 | 80,09 | 79,42 | 80,38 | 79,69 | 80,67 | 79,96 | 80,95 | 80,22 | 81,23 | 80,48 | 81,51 |
| Málaga | 76,54 | 77,45 | 76,74 | 77,67 | 76,94 | 77,89 | 77,14 | 78,11 | 77,34 | 78,32 | 77,53 | 78,53 |
| Murcia | 76,88 | 78,16 | 77,04 | 78,35 | 77,20 | 78,53 | 77,36 | 78,71 | 77,51 | 78,89 | 77,66 | 79,06 |
| Navarra | 78,77 | 79,86 | 78,95 | 80,06 | 79,13 | 80,26 | 79,30 | 80,46 | 79,48 | 80,65 | 79,64 | 80,84 |
| Ourense | 76,93 | 79,43 | 77,07 | 79,62 | 77,21 | 79,80 | 77,34 | 79,98 | 77,47 | 80,16 | 77,59 | 80,33 |
| Asturias | 76,55 | 77,57 | 76,74 | 77,78 | 76,93 | 77,98 | 77,12 | 78,19 | 77,30 | 78,39 | 77,48 | 78,59 |
| Palencia | 77,05 | 79,39 | 77,23 | 79,61 | 77,40 | 79,83 | 77,57 | 80,04 | 77,74 | 80,26 | 77,90 | 80,47 |
| Palmas, Las | 76,05 | 77,82 | 76,27 | 78,06 | 76,48 | 78,31 | 76,69 | 78,55 | 76,89 | 78,79 | 77,09 | 79,03 |
| Pontevedra | 77,35 | 78,52 | 77,58 | 78,76 | 77,80 | 79,01 | 78,02 | 79,25 | 78,24 | 79,49 | 78,45 | 79,73 |
| Salamanca | 78,82 | 81,06 | 79,00 | 81,28 | 79,17 | 81,49 | 79,34 | 81,70 | 79,50 | 81,91 | 79,66 | 82,12 |
| Santa Cruz de Tenerife | 76,48 | 78,41 | 76,64 | 78,60 | 76,79 | 78,78 | 76,93 | 78,96 | 77,08 | 79,15 | 77,21 | 79,33 |
| Cantabria | 77,50 | 78,62 | 77,68 | 78,83 | 77,87 | 79,03 | 78,05 | 79,23 | 78,22 | 79,43 | 78,40 | 79,63 |
| Segovia | 78,46 | 80,91 | 78,56 | 81,05 | 78,65 | 81,18 | 78,74 | 81,32 | 78,83 | 81,46 | 78,91 | 81,59 |
| Sevilla | 75,99 | 76,89 | 76,17 | 77,08 | 76,34 | 77,27 | 76,51 | 77,45 | 76,67 | 77,63 | 76,83 | 77,81 |
| Soria | 77,75 | 82,13 | 77,83 | 82,29 | 77,91 | 82,44 | 77,98 | 82,59 | 78,05 | 82,75 | 78,11 | 82,90 |
| Tarragona | 77,46 | 78,51 | 77,60 | 78,67 | 77,74 | 78,83 | 77,87 | 78,98 | 78,00 | 79,13 | 78,13 | 79,28 |
| Teruel | 77,70 | 80,89 | 77,79 | 81,04 | 77,87 | 81,18 | 77,96 | 81,32 | 78,03 | 81,46 | 78,11 | 81,61 |
| Toledo | 78,32 | 79,70 | 78,47 | 79,86 | 78,61 | 80,03 | 78,75 | 80,19 | 78,88 | 80,36 | 79,01 | 80,52 |
| Valencia/València | 76,92 | 78,03 | 77,14 | 78,26 | 77,35 | 78,49 | 77,56 | 78,72 | 77,76 | 78,95 | 77,96 | 79,17 |
| Valladolid | 78,19 | 79,37 | 78,32 | 79,52 | 78,46 | 79,68 | 78,59 | 79,83 | 78,72 | 79,98 | 78,84 | 80,13 |
| Bizkaia | 77,83 | 78,89 | 78,08 | 79,16 | 78,33 | 79,43 | 78,57 | 79,69 | 78,81 | 79,95 | 79,05 | 80,21 |
| Zamora | 78,30 | 81,38 | 78,46 | 81,59 | 78,61 | 81,80 | 78,76 | 82,01 | 78,91 | 82,22 | 79,05 | 82,42 |
| Zaragoza | 77,44 | 78,93 | 77,58 | 79,10 | 77,72 | 79,26 | 77,85 | 79,43 | 77,98 | 79,59 | 78,11 | 79,75 |
| Ceuta | 74,80 | 78,42 | 74,98 | 78,65 | 75,15 | 78,89 | 75,32 | 79,13 | 75,48 | 79,37 | 75,64 | 79,61 |
| Melilla | 73,56 | 79,06 | 73,66 | 79,25 | 73,76 | 79,45 | 73,85 | 79,64 | 73,94 | 79,84 | 74,01 | 80,04 |

Fuente: Elaboración propia a partir de datos del INE

La esperanza de vida al nacimiento de cada provincia tiene un intervalo de confianza centrado en el valor de predicción

$$e(t_0, x, h) = e(t_0, x) + \hat{a}(x, h) + \hat{b}(x, h) \cdot t_0 \quad (5.2)$$

Los intervalos de predicción tienen posiciones relativas análogas a la tendencia del periodo observado. Madrid y Navarra aparecen en la parte superior de la nube de intervalos de confianza mientras que Ceuta y Melilla se localizan en la parte inferior.

Tabla 5.6. Extremos del intervalo de confianza provincial al 95% proyectado en el corto plazo para la esperanza de vida al nacimiento. Varones. Periodo 2015 al 2019. Extremos predichos para el intervalo provincial al 95% de la esperanza de vida al nacimiento

| Provincias | 2015 | | 2016 | | 2017 | | 2018 | | 2019 | |
|------------------------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
| | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. | E. Inf. | E.Sup. |
| Araba/Álava | 79,44 | 81,85 | 79,62 | 82,08 | 79,80 | 82,32 | 79,97 | 82,55 | 80,15 | 82,78 |
| Albacete | 79,55 | 81,19 | 79,71 | 81,38 | 79,85 | 81,56 | 80,00 | 81,74 | 80,14 | 81,93 |
| Alicante/Alacant | 78,31 | 79,69 | 78,46 | 79,88 | 78,62 | 80,06 | 78,77 | 80,25 | 78,92 | 80,43 |
| Almería | 76,10 | 77,67 | 76,18 | 77,78 | 76,26 | 77,89 | 76,34 | 78,00 | 76,41 | 78,11 |
| Ávila | 77,98 | 80,49 | 78,03 | 80,59 | 78,08 | 80,70 | 78,12 | 80,80 | 78,17 | 80,90 |
| Badajoz | 77,87 | 79,12 | 78,03 | 79,31 | 78,20 | 79,49 | 78,35 | 79,68 | 78,51 | 79,86 |
| Balears, Illes | 79,78 | 81,49 | 80,05 | 81,80 | 80,32 | 82,11 | 80,59 | 82,41 | 80,85 | 82,72 |
| Barcelona | 79,79 | 80,37 | 80,01 | 80,60 | 80,23 | 80,84 | 80,45 | 81,07 | 80,66 | 81,29 |
| Burgos | 79,28 | 81,49 | 79,44 | 81,69 | 79,60 | 81,90 | 79,75 | 82,10 | 79,90 | 82,30 |
| Cáceres | 78,13 | 79,32 | 78,24 | 79,45 | 78,34 | 79,58 | 78,45 | 79,72 | 78,55 | 79,84 |
| Cádiz | 76,87 | 79,22 | 77,07 | 79,47 | 77,26 | 79,71 | 77,45 | 79,96 | 77,64 | 80,20 |
| Castellón/Castelló | 77,66 | 79,60 | 77,78 | 79,77 | 77,90 | 79,93 | 78,01 | 80,09 | 78,12 | 80,24 |
| Ciudad Real | 78,36 | 80,40 | 78,51 | 80,59 | 78,65 | 80,79 | 78,80 | 80,98 | 78,94 | 81,17 |
| Córdoba | 77,72 | 79,25 | 77,86 | 79,42 | 78,00 | 79,59 | 78,13 | 79,76 | 78,26 | 79,92 |
| Coruña, A | 78,20 | 79,85 | 78,39 | 80,07 | 78,58 | 80,29 | 78,76 | 80,52 | 78,94 | 80,74 |
| Cuenca | 79,33 | 81,00 | 79,41 | 81,11 | 79,48 | 81,22 | 79,55 | 81,32 | 79,62 | 81,43 |
| Girona | 79,38 | 80,39 | 79,56 | 80,59 | 79,73 | 80,79 | 79,91 | 80,99 | 80,08 | 81,18 |
| Granada | 77,42 | 78,64 | 77,55 | 78,80 | 77,68 | 78,95 | 77,80 | 79,10 | 77,92 | 79,26 |
| Guadalajara | 79,43 | 82,92 | 79,51 | 83,08 | 79,59 | 83,23 | 79,67 | 83,39 | 79,74 | 83,55 |
| Gipuzkoa | 79,48 | 81,02 | 79,71 | 81,28 | 79,94 | 81,54 | 80,16 | 81,80 | 80,38 | 82,06 |
| Huelva | 76,93 | 77,92 | 77,06 | 78,08 | 77,20 | 78,23 | 77,33 | 78,38 | 77,45 | 78,53 |
| Huesca | 77,78 | 81,24 | 77,83 | 81,36 | 77,88 | 81,48 | 77,92 | 81,59 | 77,95 | 81,71 |
| Jaén | 76,97 | 78,67 | 77,06 | 78,79 | 77,14 | 78,91 | 77,22 | 79,03 | 77,30 | 79,15 |
| León | 78,05 | 79,90 | 78,15 | 80,03 | 78,24 | 80,16 | 78,32 | 80,29 | 78,41 | 80,42 |
| Lleida | 78,53 | 80,26 | 78,67 | 80,44 | 78,81 | 80,62 | 78,95 | 80,79 | 79,09 | 80,97 |
| Rioja, La | 79,16 | 80,63 | 79,31 | 80,81 | 79,46 | 81,00 | 79,60 | 81,18 | 79,75 | 81,35 |
| Lugo | 78,36 | 80,19 | 78,55 | 80,40 | 78,72 | 80,62 | 78,90 | 80,84 | 79,07 | 81,05 |
| Madrid | 80,74 | 81,79 | 80,99 | 82,06 | 81,24 | 82,34 | 81,49 | 82,61 | 81,74 | 82,88 |
| Málaga | 77,72 | 78,74 | 77,91 | 78,95 | 78,09 | 79,16 | 78,27 | 79,36 | 78,45 | 79,57 |
| Murcia | 77,81 | 79,24 | 77,95 | 79,41 | 78,09 | 79,58 | 78,22 | 79,75 | 78,36 | 79,92 |
| Navarra | 79,81 | 81,04 | 79,97 | 81,22 | 80,13 | 81,41 | 80,29 | 81,60 | 80,45 | 81,78 |
| Ourense | 77,71 | 80,51 | 77,83 | 80,69 | 77,94 | 80,86 | 78,05 | 81,03 | 78,16 | 81,21 |
| Asturias | 77,65 | 78,79 | 77,82 | 78,98 | 77,99 | 79,18 | 78,16 | 79,37 | 78,33 | 79,56 |
| Palencia | 78,06 | 80,68 | 78,21 | 80,89 | 78,37 | 81,10 | 78,51 | 81,31 | 78,66 | 81,52 |
| Palmas, Las | 77,29 | 79,26 | 77,48 | 79,50 | 77,67 | 79,73 | 77,86 | 79,96 | 78,04 | 80,20 |
| Pontevedra | 78,66 | 79,96 | 78,87 | 80,20 | 79,07 | 80,43 | 79,27 | 80,66 | 79,47 | 80,89 |
| Salamanca | 79,82 | 82,32 | 79,97 | 82,53 | 80,12 | 82,73 | 80,26 | 82,93 | 80,41 | 83,14 |
| Santa Cruz de Tenerife | 77,35 | 79,51 | 77,48 | 79,68 | 77,61 | 79,86 | 77,74 | 80,03 | 77,86 | 80,21 |
| Cantabria | 78,57 | 79,83 | 78,73 | 80,02 | 78,90 | 80,21 | 79,06 | 80,40 | 79,22 | 80,59 |
| Segovia | 78,99 | 81,72 | 79,07 | 81,86 | 79,14 | 81,99 | 79,20 | 82,12 | 79,27 | 82,25 |
| Sevilla | 76,99 | 77,99 | 77,14 | 78,17 | 77,29 | 78,34 | 77,44 | 78,51 | 77,59 | 78,68 |
| Soria | 78,16 | 83,06 | 78,21 | 83,21 | 78,26 | 83,37 | 78,30 | 83,52 | 78,34 | 83,67 |
| Tarragona | 78,25 | 79,43 | 78,37 | 79,58 | 78,49 | 79,72 | 78,61 | 79,86 | 78,72 | 80,00 |
| Teruel | 78,17 | 81,75 | 78,24 | 81,89 | 78,30 | 82,03 | 78,36 | 82,16 | 78,41 | 82,30 |
| Toledo | 79,14 | 80,67 | 79,26 | 80,83 | 79,38 | 80,98 | 79,50 | 81,14 | 79,62 | 81,29 |
| Valencia/València | 78,16 | 79,39 | 78,35 | 79,62 | 78,55 | 79,83 | 78,74 | 80,05 | 78,92 | 80,27 |
| Valladolid | 78,96 | 80,28 | 79,08 | 80,43 | 79,19 | 80,57 | 79,31 | 80,71 | 79,41 | 80,85 |
| Bizkaia | 79,28 | 80,47 | 79,52 | 80,73 | 79,74 | 80,98 | 79,97 | 81,23 | 80,19 | 81,49 |
| Zamora | 79,19 | 82,63 | 79,32 | 82,83 | 79,45 | 83,04 | 79,57 | 83,24 | 79,69 | 83,44 |
| Zaragoza | 78,24 | 79,91 | 78,36 | 80,06 | 78,48 | 80,22 | 78,59 | 80,37 | 78,71 | 80,52 |
| Ceuta | 75,80 | 79,84 | 75,95 | 80,08 | 76,10 | 80,32 | 76,24 | 80,55 | 76,38 | 80,79 |
| Melilla | 74,09 | 80,24 | 74,16 | 80,44 | 74,22 | 80,63 | 74,28 | 80,83 | 74,34 | 81,03 |

Fuente: Elaboración propia a partir de datos del INE

La Tabla 5.7 presenta el rango o diferencia absoluta entre los extremos del intervalo de confianza al 95% de cada provincia. Se observa que la longitud de los rangos presenta una alta correlación inversa con el tamaño poblacional de la provincia, como cabía esperar. También se observa que la amplitud del rango crece con el tiempo en una tendencia francamente estable que en ningún momento es exponencial. Esta propiedad da mayor confianza en el modelo pues establece un intervalo de confianza creíble y asumible, siendo de una longitud razonablemente pequeña.

Tabla 5.7. Rango del intervalo de confianza provincial al 95% proyectado en el corto plazo para la esperanza de vida al nacimiento. Varones. Periodo 2009 al 2019.
Rango del IC proyectado para la esperanza de vida al nacimiento provincial. Varones

| Provincias | Años | | | | | | | | | | |
|------------------------|------|------|------|------|------|------|------|------|------|------|------|
| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| Araba/Alava | 2,16 | 2,20 | 2,23 | 2,28 | 2,32 | 2,37 | 2,41 | 2,46 | 2,52 | 2,57 | 2,63 |
| Albacete | 1,46 | 1,49 | 1,51 | 1,54 | 1,57 | 1,60 | 1,63 | 1,67 | 1,70 | 1,74 | 1,78 |
| Alicante/Alacant | 1,24 | 1,26 | 1,28 | 1,30 | 1,33 | 1,35 | 1,38 | 1,41 | 1,44 | 1,47 | 1,51 |
| Almería | 1,40 | 1,42 | 1,45 | 1,47 | 1,50 | 1,53 | 1,56 | 1,60 | 1,63 | 1,67 | 1,70 |
| Avila | 2,25 | 2,28 | 2,32 | 2,37 | 2,41 | 2,46 | 2,51 | 2,56 | 2,62 | 2,68 | 2,74 |
| Badajoz | 1,11 | 1,13 | 1,15 | 1,17 | 1,20 | 1,22 | 1,25 | 1,27 | 1,30 | 1,33 | 1,36 |
| Balears, Illes | 1,53 | 1,56 | 1,58 | 1,61 | 1,64 | 1,68 | 1,71 | 1,75 | 1,78 | 1,82 | 1,86 |
| Barcelona | 0,52 | 0,53 | 0,54 | 0,55 | 0,56 | 0,57 | 0,58 | 0,59 | 0,61 | 0,62 | 0,63 |
| Burgos | 1,97 | 2,01 | 2,04 | 2,08 | 2,12 | 2,16 | 2,21 | 2,25 | 2,30 | 2,35 | 2,40 |
| Cáceres | 1,07 | 1,08 | 1,10 | 1,12 | 1,14 | 1,17 | 1,19 | 1,22 | 1,24 | 1,27 | 1,30 |
| Cádiz | 2,10 | 2,14 | 2,18 | 2,22 | 2,26 | 2,30 | 2,35 | 2,40 | 2,45 | 2,51 | 2,56 |
| Castellón/Castelló | 1,74 | 1,77 | 1,80 | 1,83 | 1,87 | 1,91 | 1,95 | 1,99 | 2,03 | 2,07 | 2,12 |
| Ciudad Real | 1,83 | 1,86 | 1,89 | 1,93 | 1,96 | 2,00 | 2,04 | 2,09 | 2,13 | 2,18 | 2,23 |
| Córdoba | 1,36 | 1,39 | 1,41 | 1,44 | 1,46 | 1,49 | 1,52 | 1,56 | 1,59 | 1,62 | 1,66 |
| Coruña, A | 1,47 | 1,50 | 1,52 | 1,55 | 1,58 | 1,61 | 1,65 | 1,68 | 1,72 | 1,76 | 1,79 |
| Cuenca | 1,49 | 1,51 | 1,54 | 1,57 | 1,60 | 1,63 | 1,66 | 1,70 | 1,74 | 1,77 | 1,81 |
| Girona | 0,91 | 0,92 | 0,94 | 0,95 | 0,97 | 0,99 | 1,01 | 1,03 | 1,06 | 1,08 | 1,10 |
| Granada | 1,09 | 1,11 | 1,13 | 1,15 | 1,18 | 1,20 | 1,22 | 1,25 | 1,28 | 1,30 | 1,33 |
| Guadalajara | 3,12 | 3,18 | 3,23 | 3,29 | 3,36 | 3,42 | 3,49 | 3,57 | 3,64 | 3,72 | 3,80 |
| Gipuzkoa | 1,38 | 1,40 | 1,42 | 1,45 | 1,48 | 1,51 | 1,54 | 1,57 | 1,60 | 1,64 | 1,68 |
| Huelva | 0,89 | 0,90 | 0,92 | 0,94 | 0,95 | 0,97 | 0,99 | 1,01 | 1,04 | 1,06 | 1,08 |
| Huesca | 3,09 | 3,14 | 3,19 | 3,25 | 3,32 | 3,38 | 3,45 | 3,52 | 3,60 | 3,68 | 3,76 |
| Jaén | 1,51 | 1,54 | 1,57 | 1,60 | 1,63 | 1,66 | 1,69 | 1,73 | 1,77 | 1,80 | 1,84 |
| León | 1,65 | 1,68 | 1,71 | 1,74 | 1,77 | 1,81 | 1,85 | 1,88 | 1,92 | 1,97 | 2,01 |
| Lleida | 1,54 | 1,57 | 1,60 | 1,63 | 1,66 | 1,69 | 1,73 | 1,76 | 1,80 | 1,84 | 1,88 |
| Rioja, La | 1,32 | 1,34 | 1,37 | 1,39 | 1,42 | 1,45 | 1,48 | 1,51 | 1,54 | 1,57 | 1,61 |
| Lugo | 1,63 | 1,66 | 1,68 | 1,72 | 1,75 | 1,78 | 1,82 | 1,86 | 1,90 | 1,94 | 1,98 |
| Madrid | 0,94 | 0,96 | 0,97 | 0,99 | 1,01 | 1,03 | 1,05 | 1,07 | 1,10 | 1,12 | 1,14 |
| Málaga | 0,92 | 0,93 | 0,95 | 0,97 | 0,98 | 1,00 | 1,02 | 1,05 | 1,07 | 1,09 | 1,12 |
| Murcia | 1,28 | 1,30 | 1,32 | 1,35 | 1,37 | 1,40 | 1,43 | 1,46 | 1,49 | 1,53 | 1,56 |
| Navarra | 1,10 | 1,11 | 1,13 | 1,15 | 1,18 | 1,20 | 1,22 | 1,25 | 1,28 | 1,31 | 1,33 |
| Ourense | 2,50 | 2,54 | 2,59 | 2,63 | 2,68 | 2,74 | 2,79 | 2,85 | 2,91 | 2,98 | 3,04 |
| Asturias | 1,02 | 1,03 | 1,05 | 1,07 | 1,09 | 1,11 | 1,14 | 1,16 | 1,19 | 1,21 | 1,24 |
| Palencia | 2,35 | 2,39 | 2,43 | 2,47 | 2,52 | 2,57 | 2,62 | 2,68 | 2,74 | 2,80 | 2,86 |
| Palmas, Las | 1,77 | 1,80 | 1,83 | 1,86 | 1,90 | 1,94 | 1,98 | 2,02 | 2,06 | 2,11 | 2,15 |
| Pontevedra | 1,17 | 1,19 | 1,21 | 1,23 | 1,25 | 1,28 | 1,30 | 1,33 | 1,36 | 1,39 | 1,42 |
| Salamanca | 2,24 | 2,28 | 2,32 | 2,36 | 2,41 | 2,46 | 2,51 | 2,56 | 2,61 | 2,67 | 2,73 |
| Santa Cruz de Tenerife | 1,93 | 1,96 | 2,00 | 2,03 | 2,07 | 2,11 | 2,16 | 2,20 | 2,25 | 2,30 | 2,35 |
| Cantabria | 1,13 | 1,15 | 1,17 | 1,19 | 1,21 | 1,24 | 1,26 | 1,29 | 1,32 | 1,34 | 1,37 |
| Segovia | 2,44 | 2,49 | 2,53 | 2,58 | 2,63 | 2,68 | 2,73 | 2,79 | 2,85 | 2,91 | 2,98 |
| Sevilla | 0,90 | 0,91 | 0,93 | 0,94 | 0,96 | 0,98 | 1,00 | 1,02 | 1,05 | 1,07 | 1,09 |
| Soria | 4,38 | 4,45 | 4,53 | 4,61 | 4,70 | 4,80 | 4,90 | 5,00 | 5,11 | 5,22 | 5,33 |
| Tarragona | 1,05 | 1,07 | 1,09 | 1,11 | 1,13 | 1,15 | 1,18 | 1,20 | 1,23 | 1,26 | 1,28 |
| Teruel | 3,20 | 3,25 | 3,31 | 3,37 | 3,43 | 3,50 | 3,57 | 3,65 | 3,73 | 3,81 | 3,89 |
| Toledo | 1,37 | 1,40 | 1,42 | 1,45 | 1,47 | 1,50 | 1,53 | 1,57 | 1,60 | 1,64 | 1,67 |
| Valencia/València | 1,10 | 1,12 | 1,14 | 1,16 | 1,19 | 1,21 | 1,24 | 1,26 | 1,29 | 1,32 | 1,35 |
| Valladolid | 1,18 | 1,20 | 1,22 | 1,24 | 1,27 | 1,29 | 1,32 | 1,35 | 1,38 | 1,41 | 1,44 |
| Bizkaia | 1,06 | 1,08 | 1,10 | 1,12 | 1,14 | 1,16 | 1,19 | 1,21 | 1,24 | 1,26 | 1,29 |
| Zamora | 3,08 | 3,13 | 3,19 | 3,24 | 3,31 | 3,37 | 3,44 | 3,52 | 3,59 | 3,67 | 3,75 |
| Zaragoza | 1,49 | 1,52 | 1,55 | 1,57 | 1,60 | 1,64 | 1,67 | 1,71 | 1,74 | 1,78 | 1,82 |
| Ceuta | 3,62 | 3,68 | 3,74 | 3,81 | 3,89 | 3,96 | 4,05 | 4,13 | 4,22 | 4,31 | 4,41 |
| Melilla | 5,50 | 5,59 | 5,69 | 5,79 | 5,91 | 6,02 | 6,15 | 6,28 | 6,41 | 6,55 | 6,70 |

Fuente: Elaboración propia a partir de datos del INE

El rango más pequeño en 2009 es para Barcelona con una longitud de 0,52 años frente a los 5,5 años de la longitud del rango de Melilla, la población de menor tamaño. La regla en cuanto al tamaño poblacional no es fija, pues Madrid no es el mínimo aunque sí de los más pequeños. En el otro extremo tampoco Ceuta es el segundo mayor, superado por Soria cuya población es la tercera de menor tamaño.

Análogamente al análisis realizado con los datos observados del periodo 2009-2011, a continuación se estudia el éxito del intervalo de confianza proporcionado por la proyección a corto plazo proporcionado por el INE a lo largo del periodo 2009 hasta 2019 con los datos derivados de las proyecciones de población a corto plazo 2010-2020.

Para cada año y edad, la fórmula 5.1 determina un intervalo de confianza. Dado un nivel de confianza del 95% se tienen 109.604³ intervalos de confianza que testar. Se define el éxito como que la esperanza de vida observada cae dentro del intervalo de confianza predicho por la fórmula 5.1 a partir de la proyección de la esperanza de vida al nacimiento nacional para los 11 años de prueba, las 52 provincias, los 2 sexos y las edades. El recuento total de éxitos es 106.104, que supone un 96,81%, de todos los casos.

Por sexos, como se ve en la Tabla 5.8, el porcentaje es superior para mujeres con un 97,84% frente a los varones con un 95,77%. Por años el porcentaje decrece pasando de 99,50% en 2009 al 92,97% en 2019.

Tabla 5.8. Porcentaje de éxito para el intervalo de confianza provincial por sexo y año.

Porcentaje de éxito para el Intervalo de Confianza provincial por sexo y año al 95%

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Total |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Varones | 99,60% | 99,56% | 99,00% | 97,91% | 97,19% | 96,35% | 95,70% | 94,20% | 92,63% | 91,21% | 90,16% | 95,77% |
| Mujeres | 99,40% | 99,20% | 99,00% | 98,88% | 98,55% | 98,11% | 97,57% | 97,07% | 96,55% | 96,13% | 95,78% | 97,84% |
| Total | 99,50% | 99,38% | 99,00% | 98,39% | 97,87% | 97,23% | 96,64% | 95,63% | 94,59% | 93,67% | 92,97% | 96,81% |

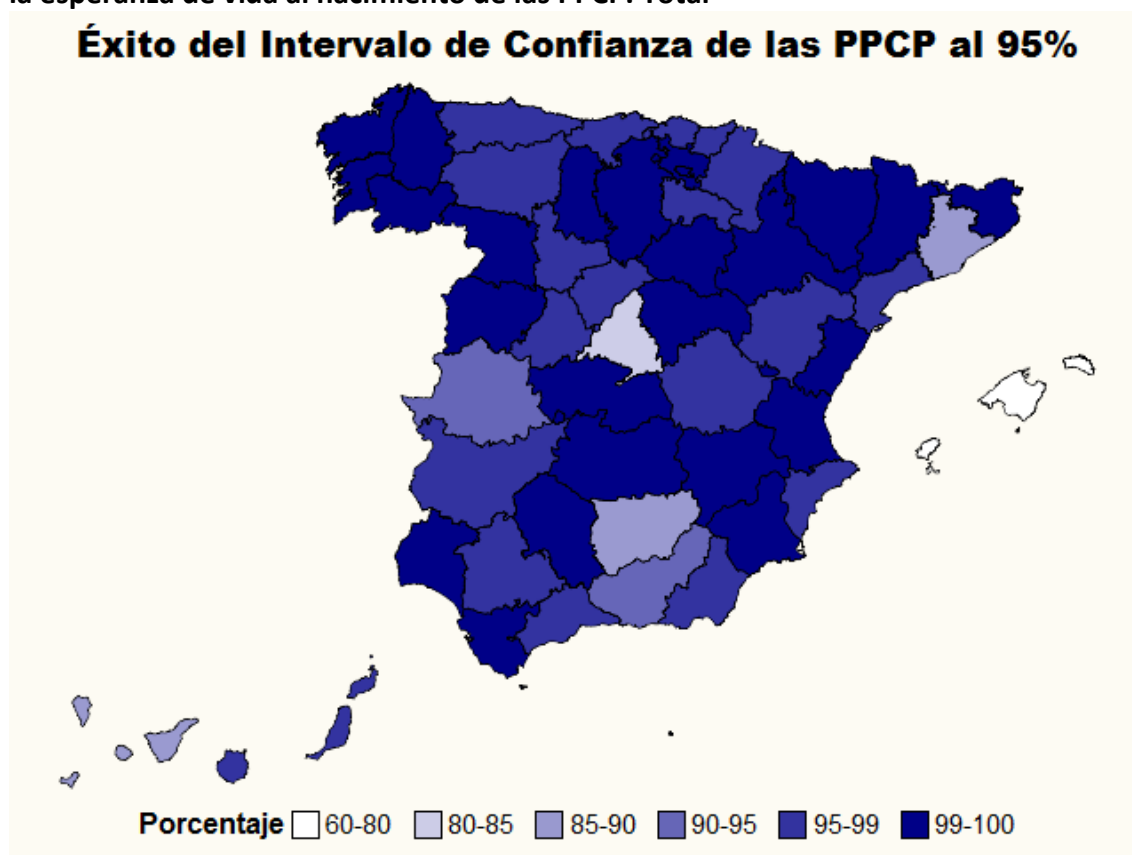
Fuente: Elaboración propia a partir de datos del INE

Merecen especial atención los resultados que se obtienen hasta el año 2010, un porcentaje de acierto del 99,38%. Este año es el actual desde la perspectiva de la simulación y proporciona los parámetros que luego se introducen para producir las *Estimaciones de Población Actual*.

La distribución geográfica del porcentaje de éxito es aleatoria como se puede ver en la Figura 5.3. No es posible distinguir regiones aunque sí, en general, un porcentaje alto a excepción de unos pocos casos. Además, se parece poco a la distribución del éxito de la validación con los datos observados.

³ 11 años x 2 sexos x 96 edades x 50 provincias + 11 años 2 sexos x 91 edades x 2 Ciudades Autónomas.

Figura 5.3. Porcentaje de éxito del intervalo de confianza al 95% por provincias para la esperanza de vida al nacimiento de las PPCP. Total



Fuente: Elaboración propia a partir de datos del INE

En la Tabla 5.9 se observa que casi todas las provincias tienen porcentajes de acierto que superan ampliamente el 85%. La proporción es independiente del tamaño de la provincia. Por otra parte, las provincias de pequeño tamaño como Ceuta o Melilla tienen un porcentaje alto de éxito. Hay 45 provincias que superan el 95% y la mediana del porcentaje de éxito es del 98,91%.

**Tabla 5.9. Porcentaje de éxito del intervalo de confianza por provincia.
Porcentaje de éxito para el IC por provincias al 95% de la PPCP 2010-2020**

| Provincias | Varones | Mujeres | Total | Provincias | Varones | Mujeres | Total |
|--------------------|---------|---------|---------|------------------------|---------------|---------------|---------------|
| Araba/Álava | 100,00% | 98,30% | 99,15% | Navarra | 100,00% | 97,35% | 98,67% |
| Albacete | 100,00% | 99,62% | 99,81% | Ourense | 98,48% | 100,00% | 99,24% |
| Alicante/Alacant | 97,06% | 99,24% | 98,15% | Asturias | 99,72% | 93,94% | 96,83% |
| Almería | 95,55% | 100,00% | 97,77% | Palencia | 100,00% | 99,91% | 99,95% |
| Ávila | 91,95% | 100,00% | 95,98% | Palmas, Las | 95,64% | 99,15% | 97,40% |
| Badajoz | 100,00% | 94,51% | 97,25% | Pontevedra | 100,00% | 100,00% | 100,00% |
| Balears, Illes | 50,85% | 74,05% | 62,45% | Salamanca | 100,00% | 100,00% | 100,00% |
| Barcelona | 75,66% | 98,86% | 87,26% | Santa Cruz de Tenerife | 81,34% | 89,58% | 85,46% |
| Burgos | 100,00% | 100,00% | 100,00% | Cantabria | 97,73% | 100,00% | 98,86% |
| Cáceres | 85,04% | 100,00% | 92,52% | Segovia | 90,06% | 100,00% | 95,03% |
| Cádiz | 100,00% | 100,00% | 100,00% | Sevilla | 100,00% | 95,17% | 97,59% |
| Castellón/Castelló | 100,00% | 100,00% | 100,00% | Soria | 100,00% | 100,00% | 100,00% |
| Ciudad Real | 100,00% | 100,00% | 100,00% | Tarragona | 97,25% | 100,00% | 98,63% |
| Córdoba | 100,00% | 100,00% | 100,00% | Teruel | 100,00% | 97,92% | 98,96% |
| Coruña, A | 100,00% | 100,00% | 100,00% | Toledo | 100,00% | 100,00% | 100,00% |
| Cuenca | 98,67% | 93,18% | 95,93% | Valencia/València | 99,15% | 100,00% | 99,57% |
| Girona | 100,00% | 100,00% | 100,00% | Valladolid | 100,00% | 96,02% | 98,01% |
| Granada | 95,74% | 93,84% | 94,79% | Bizkaia | 98,96% | 95,83% | 97,40% |
| Guadalajara | 100,00% | 100,00% | 100,00% | Zamora | 100,00% | 99,05% | 99,53% |
| Gipuzkoa | 96,40% | 98,96% | 97,68% | Zaragoza | 98,96% | 100,00% | 99,48% |
| Huelva | 99,34% | 100,00% | 99,67% | Ceuta | 100,00% | 100,00% | 100,00% |
| Huesca | 99,43% | 100,00% | 99,72% | Melilla | 100,00% | 92,71% | 96,35% |
| Jaén | 89,77% | 86,74% | 88,26% | Total | 95,77% | 97,84% | 96,81% |
| León | 90,91% | 100,00% | 95,45% | | | | |
| Lleida | 100,00% | 100,00% | 100,00% | | | | |
| Rioja, La | 100,00% | 93,94% | 96,97% | | | | |
| Lugo | 99,72% | 100,00% | 99,86% | | | | |
| Madrid | 62,22% | 100,00% | 81,11% | | | | |
| Málaga | 95,08% | 99,62% | 97,35% | | | | |
| Murcia | 100,00% | 100,00% | 100,00% | | | | |

Fuente: Elaboración propia a partir de datos del INE

El porcentaje de éxito es alto, salvo para dos provincias, Baleares que repite mínimo con un 62,45% del total y un caso particular de 50,85% en varones y Madrid con un 81,11% aunque debido exclusivamente a los varones con un 62,22% pues las mujeres presentan un porcentaje de éxito del 100% en esta provincia.

En la Tabla 5.10, Tabla 5.11 y la Tabla 5.12 se presentan los porcentajes de éxito por provincia y año para cada sexo.

En la Tabla 5.10 de varones, todas las provincias presentan porcentajes altos de éxito salvo Baleares, cuyo porcentaje disminuye hasta un 15,63% en 2019 y Madrid que llega hasta un 33,33%, en ese último año.

Tabla 5.10. Porcentaje de éxito para el Intervalo de Confianza de la esperanza de vida al nacimiento por provincia para las Proyecciones a Corto Plazo. Varones
Porcentaje de éxito para el intervalo de confianza provincial al 95% por año. Varones

| Provincias | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Total |
|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Araba/Alava | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Albacete | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Alicante/Alacant | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 95,83% | 91,67% | 91,67% | 91,67% | 97,06% |
| Almería | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 97,92% | 90,63% | 82,29% | 80,21% | 95,55% |
| Ávila | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 90,63% | 85,42% | 69,79% | 65,63% | 91,95% |
| Badajoz | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Balears, Illes | 100,00% | 100,00% | 88,54% | 70,83% | 60,42% | 37,50% | 27,08% | 21,88% | 20,83% | 16,67% | 15,63% | 50,85% |
| Barcelona | 96,88% | 96,88% | 95,83% | 83,33% | 71,88% | 68,75% | 65,63% | 64,58% | 63,54% | 62,50% | 62,50% | 75,66% |
| Burgos | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cáceres | 100,00% | 100,00% | 100,00% | 97,92% | 95,83% | 93,75% | 91,67% | 80,21% | 64,58% | 58,33% | 53,13% | 85,04% |
| Cádiz | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Castellón/Castelló | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ciudad Real | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Córdoba | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Coruña, A | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cuenca | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 97,92% | 96,88% | 95,83% | 98,67% |
| Girona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Granada | 100,00% | 100,00% | 100,00% | 98,96% | 95,83% | 94,79% | 93,75% | 92,71% | 92,71% | 92,71% | 91,67% | 95,74% |
| Guadalajara | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Gipuzkoa | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 97,92% | 93,75% | 90,63% | 90,63% | 87,50% | 96,40% |
| Huelva | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 96,88% | 95,83% | 99,34% |
| Huesca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 93,75% | 99,43% |
| Jaén | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 96,88% | 95,83% | 82,29% | 75,00% | 71,88% | 67,71% | 89,77% |
| León | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 90,63% | 80,21% | 67,71% | 62,50% | 90,91% |
| Lleida | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Rioja, La | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Lugo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 97,92% | 99,72% |
| Madrid | 100,00% | 100,00% | 86,46% | 67,71% | 64,58% | 61,46% | 54,17% | 44,79% | 37,50% | 34,38% | 33,33% | 62,22% |
| Málaga | 98,96% | 98,96% | 97,92% | 95,83% | 93,75% | 93,75% | 93,75% | 93,75% | 93,75% | 92,71% | 92,71% | 95,08% |
| Murcia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Navarra | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ourense | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 93,75% | 91,67% | 98,48% |
| Asturias | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 99,72% |
| Palencia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Palmas, Las | 100,00% | 98,96% | 96,88% | 96,88% | 96,88% | 95,83% | 93,75% | 93,75% | 92,71% | 92,71% | 91,67% | 95,64% |
| Pontevedra | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Salamanca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Santa Cruz de Tenerife | 83,33% | 82,29% | 82,29% | 82,29% | 81,25% | 81,25% | 81,25% | 80,21% | 80,21% | 80,21% | 80,21% | 81,34% |
| Cantabria | 100,00% | 100,00% | 100,00% | 100,00% | 97,92% | 97,92% | 96,88% | 96,88% | 95,83% | 94,79% | 94,79% | 97,73% |
| Segovia | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 92,71% | 89,58% | 86,46% | 79,17% | 76,04% | 67,71% | 90,06% |
| Sevilla | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Soria | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Tarragona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 94,79% | 90,63% | 84,38% | 97,25% |
| Teruel | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Toledo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Valencia/València | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 97,92% | 97,92% | 97,92% | 99,15% |
| Valladolid | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Bizkaia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 97,92% | 96,88% | 96,88% | 96,88% | 98,96% |
| Zamora | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Zaragoza | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 97,92% | 97,92% | 97,92% | 97,92% | 97,92% | 98,96% |
| Ceuta | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Melilla | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Total | 99,60% | 99,56% | 99,00% | 97,91% | 97,19% | 96,35% | 95,70% | 94,20% | 92,63% | 91,21% | 90,16% | 95,77% |

Fuente: Elaboración propia a partir de datos del INE

De nuevo merecen especial atención los resultados obtenidos hasta el año 2010. Años en los que se calculan los parámetros que estarán implicados en la *Estimación de Población Actual*.

En el caso de los varones, el año 2010 alcanza un porcentaje total de acierto del 99,56%. Las únicas provincias en las que el éxito no es total son Barcelona con un 96,88%, Málaga y Las Palmas con un 98,96% y Santa Cruz de Tenerife con un 82,33%.

A continuación, en la Tabla 5.11, se detallan los resultados para mujeres.

Estudio de la Mortalidad en España y sus Provincias

Tabla 5.11. Porcentaje de éxito para el Intervalo de Confianza de la esperanza de vida al nacimiento por provincia para las Proyecciones a Corto Plazo. Mujeres
Porcentaje de éxito para el intervalo de confianza provincial al 95% por año. Mujeres

| Provincias | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Total |
|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Araba/Alava | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 96,88% | 95,83% | 95,83% | 93,75% | 98,30% |
| Albacete | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 99,62% |
| Alicante/Alacant | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 99,24% |
| Almería | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ávila | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Badajoz | 98,96% | 97,92% | 95,83% | 95,83% | 95,83% | 95,83% | 92,71% | 91,67% | 91,67% | 91,67% | 91,67% | 94,51% |
| Balears, Illes | 100,00% | 100,00% | 100,00% | 100,00% | 88,54% | 75,00% | 58,33% | 57,29% | 52,08% | 43,75% | 39,58% | 74,05% |
| Barcelona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 96,88% | 96,88% | 98,86% |
| Burgos | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cáceres | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cádiz | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Castellón/Castelló | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ciudad Real | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Córdoba | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Coruña, A | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cuenca | 100,00% | 96,88% | 96,88% | 94,79% | 93,75% | 91,67% | 91,67% | 90,63% | 89,58% | 89,58% | 89,58% | 93,18% |
| Girona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Granada | 95,83% | 95,83% | 95,83% | 94,79% | 94,79% | 94,79% | 94,79% | 93,75% | 90,63% | 90,63% | 90,63% | 93,84% |
| Guadalajara | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Gipuzkoa | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 97,92% | 97,92% | 97,92% | 97,92% | 97,92% | 98,96% |
| Huelva | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Huesca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Jaén | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 96,88% | 93,75% | 79,17% | 68,75% | 60,42% | 55,21% | 86,74% |
| León | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Lleida | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Rioja, La | 97,92% | 97,92% | 95,83% | 94,79% | 93,75% | 93,75% | 92,71% | 91,67% | 91,67% | 91,67% | 91,67% | 93,94% |
| Lugo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Madrid | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Málaga | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,96% | 98,96% | 99,62% |
| Murcia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Navarra | 98,96% | 97,92% | 97,92% | 97,92% | 97,92% | 97,92% | 97,92% | 96,88% | 96,88% | 95,83% | 94,79% | 97,35% |
| Ourense | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Asturias | 95,83% | 95,83% | 95,83% | 93,75% | 93,75% | 93,75% | 92,71% | 92,71% | 92,71% | 92,71% | 92,71% | 93,94% |
| Palencia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 99,91% |
| Palmas, Las | 96,88% | 96,88% | 96,88% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,15% |
| Pontevedra | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Salamanca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Santa Cruz de Tenerife | 93,75% | 90,63% | 89,58% | 89,58% | 89,58% | 89,58% | 88,54% | 88,54% | 88,54% | 88,54% | 88,54% | 89,58% |
| Cantabria | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Segovia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Sevilla | 95,83% | 95,83% | 95,83% | 95,83% | 94,79% | 94,79% | 94,79% | 94,79% | 94,79% | 94,79% | 94,79% | 95,17% |
| Soria | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Tarragona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Teruel | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 97,92% | 97,92% | 95,83% | 93,75% | 92,71% | 97,92% |
| Toledo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Valencia/València | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Valladolid | 100,00% | 98,96% | 97,92% | 96,88% | 94,79% | 94,79% | 94,79% | 94,79% | 94,79% | 94,79% | 93,75% | 96,02% |
| Bizkaia | 96,88% | 96,88% | 95,83% | 95,83% | 95,83% | 95,83% | 95,83% | 95,83% | 95,83% | 94,79% | 94,79% | 95,83% |
| Zamora | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 97,92% | 96,88% | 94,79% | 99,05% |
| Zaragoza | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ceuta | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Melilla | 97,80% | 96,70% | 93,41% | 92,31% | 92,31% | 91,21% | 91,21% | 91,21% | 91,21% | 91,21% | 91,21% | 92,71% |
| Total | 99,40% | 99,20% | 99,00% | 98,88% | 98,55% | 98,11% | 97,57% | 97,07% | 96,55% | 96,13% | 95,78% | 97,84% |

Fuente: Elaboración propia a partir de datos del INE

De acuerdo con la Tabla 5.11 el comportamiento en las mujeres es bueno salvo en Jaén donde se aprecia un descenso rápido del porcentaje a partir del año 2014.

En el caso de las mujeres, el año 2010 alcanza un porcentaje total de acierto del 99,20%. Las únicas provincias en las que el éxito no es total son Badajoz con un 97,92%, Cuenca con un 96,88%, Granada con un 95,83%, La Rioja con un 97,92%, Navarra con un 97,92%, Asturias con un 95,83%, Las Palmas con un 96,88%, Santa

Estudio de la Mortalidad en España y sus Provincias

Cruz de Tenerife con un 90,63%, Sevilla con un 95,83%, Valladolid con un 98,96%, Vizcaya con un 96,88% y Melilla con un 96,70%.

A continuación se presenta la Tabla 5.12 con el resultado para el total de los varones y las mujeres.

Tabla 5.12. Porcentaje de éxito para el Intervalo de Confianza de la esperanza de vida al nacimiento por provincia para las Proyecciones a Corto Plazo. Total
Porcentaje de éxito para el intervalo de confianza provincial al 95% por año. Total

| Provincias | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Total |
|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Araba/Alava | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 98,44% | 97,92% | 97,92% | 96,88% | 99,15% |
| Albacete | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 99,48% | 99,48% | 99,81% |
| Alicante/Alacant | 100,00% | 100,00% | 100,00% | 99,48% | 98,96% | 98,96% | 98,96% | 97,40% | 95,31% | 95,31% | 95,31% | 98,15% |
| Almería | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 95,31% | 91,15% | 90,10% | 97,77% |
| Ávila | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 95,31% | 92,71% | 84,90% | 82,81% | 95,98% |
| Badajoz | 99,48% | 98,96% | 97,92% | 97,92% | 97,92% | 97,92% | 96,35% | 95,83% | 95,83% | 95,83% | 95,83% | 97,25% |
| Balears, Illes | 100,00% | 100,00% | 94,27% | 85,42% | 74,48% | 56,25% | 42,71% | 39,58% | 36,46% | 30,21% | 27,60% | 62,45% |
| Barcelona | 98,44% | 98,44% | 97,92% | 91,67% | 85,94% | 83,85% | 82,29% | 81,77% | 80,21% | 79,69% | 79,69% | 87,26% |
| Burgos | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cáceres | 100,00% | 100,00% | 100,00% | 98,96% | 97,92% | 96,88% | 95,83% | 90,10% | 82,29% | 79,17% | 76,56% | 92,52% |
| Cádiz | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Castellón/Castelló | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Ciudad Real | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Córdoba | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Coruña, A | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Cuenca | 100,00% | 98,44% | 98,44% | 96,88% | 96,35% | 95,31% | 95,31% | 94,79% | 93,75% | 93,23% | 92,71% | 95,93% |
| Girona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Granada | 97,92% | 97,92% | 97,92% | 96,88% | 95,31% | 94,79% | 94,27% | 93,23% | 91,67% | 91,67% | 91,15% | 94,79% |
| Guadalajara | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Gipuzkoa | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 97,92% | 95,83% | 94,27% | 94,27% | 92,71% | 97,68% |
| Huelva | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,44% | 97,92% | 99,67% |
| Huesca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 96,88% | 99,72% |
| Jaén | 100,00% | 100,00% | 100,00% | 99,48% | 99,48% | 96,88% | 94,79% | 80,73% | 71,88% | 66,15% | 61,46% | 88,26% |
| León | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 95,31% | 90,10% | 83,85% | 81,25% | 95,45% |
| Lleida | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Rioja, La | 98,96% | 98,96% | 97,92% | 97,40% | 96,88% | 96,88% | 96,35% | 95,83% | 95,83% | 95,83% | 95,83% | 96,97% |
| Lugo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 98,96% | 99,86% |
| Madrid | 100,00% | 100,00% | 93,23% | 83,85% | 82,29% | 80,73% | 77,08% | 72,40% | 68,75% | 67,19% | 66,67% | 81,11% |
| Málaga | 99,48% | 99,48% | 98,96% | 97,92% | 96,88% | 96,88% | 96,88% | 96,35% | 96,35% | 95,83% | 95,83% | 97,35% |
| Murcia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Navarra | 99,48% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 98,44% | 98,44% | 97,92% | 97,40% | 98,67% |
| Ourense | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 99,48% | 96,88% | 95,83% | 99,24% |
| Asturias | 97,92% | 97,92% | 97,92% | 96,88% | 96,88% | 96,88% | 96,88% | 96,35% | 95,83% | 95,83% | 95,83% | 96,83% |
| Palencia | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 99,95% |
| Palmas, Las | 98,44% | 97,92% | 96,88% | 98,44% | 98,44% | 97,92% | 96,88% | 96,35% | 96,35% | 95,83% | 95,83% | 97,40% |
| Pontevedra | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Salamanca | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Santa Cruz de Tenerife | 88,54% | 86,46% | 85,94% | 85,94% | 85,42% | 85,42% | 84,90% | 84,38% | 84,38% | 84,38% | 84,38% | 85,46% |
| Cantabria | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,96% | 98,44% | 98,44% | 97,92% | 97,40% | 97,40% | 98,86% |
| Segovia | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 96,35% | 94,79% | 93,23% | 89,58% | 88,02% | 83,85% | 95,03% |
| Sevilla | 97,92% | 97,92% | 97,92% | 97,92% | 97,40% | 97,40% | 97,40% | 97,40% | 97,40% | 97,40% | 97,40% | 97,59% |
| Soria | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Tarragona | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 97,40% | 95,31% | 92,19% | 98,63% |
| Teruel | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 98,96% | 98,96% | 97,92% | 96,88% | 96,35% | 98,96% |
| Toledo | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Valencia/València | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 99,48% | 99,48% | 98,96% | 98,96% | 98,96% | 99,57% |
| Valladolid | 100,00% | 99,48% | 98,96% | 98,44% | 97,40% | 97,40% | 97,40% | 97,40% | 97,40% | 97,40% | 96,88% | 98,01% |
| Bizkaia | 98,44% | 98,44% | 97,92% | 97,92% | 97,92% | 97,92% | 97,92% | 96,88% | 96,35% | 95,83% | 95,83% | 97,40% |
| Zamora | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 98,96% | 98,44% | 97,40% | 99,53% |
| Zaragoza | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 99,48% | 98,96% | 98,96% | 98,96% | 98,96% | 98,96% | 99,48% |
| Ceuta | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| Melilla | 98,90% | 98,35% | 96,70% | 96,15% | 96,15% | 95,60% | 95,60% | 95,60% | 95,60% | 95,60% | 95,60% | 96,35% |
| Total | 99,50% | 99,38% | 99,00% | 98,39% | 97,87% | 97,23% | 96,64% | 95,63% | 94,59% | 93,67% | 92,97% | 96,81% |

Fuente: Elaboración propia a partir de datos del INE

De acuerdo con la Tabla 5.12 los porcentajes de éxito están por encima del 85% y se alcanza el mínimo en Santa Cruz de Tenerife que se mantiene estable a lo largo del tiempo sin bajar del 84%.

En el año 2010 se alcanza un porcentaje de acierto para el total del 99,38%. Las únicas provincias en las que el éxito no es total son Badajoz con un 98,86%, Barcelona con un 98,44%, Cuenca con un 98,44%, Granada con un 97,92%, La Rioja con un 98,96%, Navarra con un 98,96%, Asturias con un 97,92%, Las Palmas con un 97,92%, Santa Cruz de Tenerife con un 86,46%, Sevilla con un 97,92%, Valladolid con un 99,48%, Vizcaya con un 98,44% y Melilla con un 98,35%.

La Tabla 5.13 presenta el porcentaje de éxito por edad. En general es superior al 95% en todas las edades. La serie tiene un mínimo en las primeras edades con un 95,63 a la edad de un año y luego crece paulatinamente hasta un máximo en torno a los 76 años para luego descender. En las últimas 10 edades es menor que el resto, aunque siempre por encima del 90%. Se aprecia que el porcentaje de éxito se conserva por encima del 95% y va creciendo paulatinamente hasta su máximo en los 70 y hasta los 80 años de edad donde alcanza un éxito del 100%. Para los varones el máximo se alcanza sobre los 70 y para las mujeres entre los 65 y los 80 años. Posteriormente para edades más avanzadas el porcentaje de éxito disminuye quedándose en un 90% para las últimas edades. De nuevo, como en el caso de la validación contra los datos observados se aprecia correlación con el número de defunciones teórico $d(t, x)$ como se puede ver en la Figura 3.3 del capítulo 3.

**Tabla 5.13. Porcentaje de éxito para el Intervalo de Confianza al 95% por edad.
Porcentaje de éxito por edad para un intervalo del 95%**

| Edad | Varones | Mujeres | Total | Edad | Varones | Mujeres | Total |
|------|---------|---------|--------|------|---------|---------|---------|
| 0 | 94,76% | 98,60% | 96,68% | 48 | 96,33% | 99,83% | 98,08% |
| 1 | 93,36% | 97,90% | 95,63% | 49 | 95,80% | 99,65% | 97,73% |
| 2 | 93,71% | 97,90% | 95,80% | 50 | 95,63% | 99,83% | 97,73% |
| 3 | 93,53% | 98,25% | 95,89% | 51 | 95,28% | 100,00% | 97,64% |
| 4 | 93,36% | 98,08% | 95,72% | 52 | 95,45% | 100,00% | 97,73% |
| 5 | 93,53% | 98,25% | 95,89% | 53 | 95,80% | 100,00% | 97,90% |
| 6 | 94,23% | 98,25% | 96,24% | 54 | 96,50% | 99,83% | 98,16% |
| 7 | 93,71% | 98,25% | 95,98% | 55 | 96,85% | 99,83% | 98,34% |
| 8 | 93,71% | 98,60% | 96,15% | 56 | 96,68% | 99,13% | 97,90% |
| 9 | 93,53% | 98,60% | 96,07% | 57 | 96,68% | 99,13% | 97,90% |
| 10 | 93,88% | 98,43% | 96,15% | 58 | 96,85% | 99,30% | 98,08% |
| 11 | 94,06% | 98,43% | 96,24% | 59 | 97,03% | 99,13% | 98,08% |
| 12 | 93,53% | 98,43% | 95,98% | 60 | 97,55% | 99,30% | 98,43% |
| 13 | 93,71% | 98,43% | 96,07% | 61 | 96,85% | 99,65% | 98,25% |
| 14 | 93,36% | 98,43% | 95,89% | 62 | 96,68% | 99,65% | 98,16% |
| 15 | 93,18% | 98,60% | 95,89% | 63 | 96,85% | 99,83% | 98,34% |
| 16 | 92,83% | 98,78% | 95,80% | 64 | 98,43% | 100,00% | 99,21% |
| 17 | 93,36% | 98,78% | 96,07% | 65 | 97,55% | 100,00% | 98,78% |
| 18 | 93,53% | 98,43% | 95,98% | 66 | 97,90% | 100,00% | 98,95% |
| 19 | 93,53% | 98,60% | 96,07% | 67 | 98,43% | 100,00% | 99,21% |
| 20 | 93,36% | 98,60% | 95,98% | 68 | 99,13% | 100,00% | 99,56% |
| 21 | 93,88% | 98,43% | 96,15% | 69 | 99,30% | 100,00% | 99,65% |
| 22 | 93,88% | 98,43% | 96,15% | 70 | 100,00% | 100,00% | 100,00% |
| 23 | 94,58% | 98,25% | 96,42% | 71 | 100,00% | 99,30% | 99,65% |
| 24 | 94,58% | 98,43% | 96,50% | 72 | 100,00% | 99,48% | 99,74% |
| 25 | 94,58% | 98,43% | 96,50% | 73 | 100,00% | 99,30% | 99,65% |
| 26 | 94,23% | 98,60% | 96,42% | 74 | 100,00% | 99,65% | 99,83% |
| 27 | 94,41% | 98,78% | 96,59% | 75 | 100,00% | 99,83% | 99,91% |
| 28 | 94,58% | 98,78% | 96,68% | 76 | 100,00% | 100,00% | 100,00% |
| 29 | 94,76% | 98,78% | 96,77% | 77 | 98,25% | 100,00% | 99,13% |
| 30 | 95,28% | 98,95% | 97,12% | 78 | 97,20% | 100,00% | 98,60% |
| 31 | 95,98% | 98,78% | 97,38% | 79 | 97,38% | 100,00% | 98,69% |
| 32 | 96,50% | 98,78% | 97,64% | 80 | 96,33% | 100,00% | 98,16% |
| 33 | 96,68% | 98,78% | 97,73% | 81 | 97,20% | 100,00% | 98,60% |
| 34 | 96,85% | 98,78% | 97,81% | 82 | 97,20% | 99,48% | 98,34% |
| 35 | 97,20% | 98,95% | 98,08% | 83 | 95,63% | 97,55% | 96,59% |
| 36 | 97,20% | 98,95% | 98,08% | 84 | 95,28% | 95,63% | 95,45% |
| 37 | 97,03% | 99,13% | 98,08% | 85 | 93,88% | 90,56% | 92,22% |
| 38 | 97,20% | 99,13% | 98,16% | 86 | 94,06% | 89,16% | 91,61% |
| 39 | 97,38% | 99,13% | 98,25% | 87 | 94,06% | 87,06% | 90,56% |
| 40 | 96,85% | 99,30% | 98,08% | 88 | 92,48% | 89,86% | 91,17% |
| 41 | 97,20% | 99,48% | 98,34% | 89 | 90,38% | 87,41% | 88,90% |
| 42 | 97,73% | 99,48% | 98,60% | 90 | 92,48% | 89,69% | 91,08% |
| 43 | 97,55% | 99,65% | 98,60% | 91 | 92,18% | 85,45% | 88,82% |
| 44 | 97,20% | 99,48% | 98,34% | 92 | 94,00% | 84,00% | 89,00% |
| 45 | 98,08% | 99,48% | 98,78% | 93 | 93,64% | 85,64% | 89,64% |
| 46 | 97,20% | 99,65% | 98,43% | 94 | 93,64% | 88,18% | 90,91% |
| 47 | 97,20% | 99,65% | 98,43% | 95 | 93,45% | 90,18% | 91,82% |

Fuente: Elaboración propia a partir de datos del INE

5.5 Resultados y Conclusiones

En este capítulo se ha validado el comportamiento del modelo propuesto en el apartado 5 del capítulo 4 frente a los datos observados, y luego se ha comparado con los datos de la esperanza de vida calculados para las proyecciones de población a corto plazo.

En ambos casos, el resultado ha sido satisfactorio alcanzando porcentajes de éxito por encima del 95% en la mayoría de las clasificaciones estudiadas.

En el apartado 5.2 se describen los resultados observados que servirán como elementos de contraste para validar el modelo. La esperanza de vida observada confirma la evolución de la serie desde 1991. Por una parte, existe una tendencia común de todas las provincias que se traduce en crecimiento paulatino y constante del total nacional. Por otra, se aprecia una distribución geográfica relativa que se conserva a lo largo del tiempo. La distribución es independiente del sexo en el sentido de que la ley que gobierna la mortalidad afecta por igual a las dos poblaciones sin distinción significativa ni discriminación por razones de sexo. De nuevo resalta la continuidad que se aprecia en la superficie que determinan las provincias. Este factor induce a pensar en, al menos, una dependencia por contigüidad. La extensión de este razonamiento lleva a plantear un mapa de relieve con la esperanza de vida. Este estudio queda a una línea de investigación abierta interesante pero condicionada al detalle de las fuentes de información disponibles para la población stock objeto de estudio. El futuro apunta en esta dirección gracias al desarrollo actual de los censos de población, que tienden a un seguimiento de las personas en tiempo real, de la mano de proyectos como el SIG_PT del INE, ya descrito en el apartado 2.7 de este trabajo.

Para validar los resultados del segundo apartado, en el apartado 5.3 se calculan los diferenciales de la esperanza de vida observados y se mide el porcentaje de acierto en el intervalo de confianza proyectado en el apartado 4.5. El porcentaje de éxito es notable, por encima del 90% global y similar para varones y mujeres. El resultado por año también es estable en el tiempo con un significativo valor superior

en el último año de prueba, 2011, con un valor superior al 95%. El porcentaje de éxito por provincias es en general bueno salvo para dos provincias sin motivo aparente. El porcentaje de éxito por edad es estable casi siempre por encima del 90% y con una ligera tendencia creciente hasta los 90 años.

Este primer resultado afianza la confianza en el modelo por contraste con los datos observados. Para corroborar la bondad del modelo en el futuro se compara con las proyecciones a corto plazo del INE, en el apartado 5.4. El modelo genera confianza porque en el rango de los intervalos de confianza proyectados crece con el tiempo en una tendencia estable que en ningún momento es exponencial. Como cabía esperar, la amplitud del rango depende en gran medida del tamaño de la población. Por otra parte se confirma el alto porcentaje de coincidencia con la esperanza de vida proyectada por el INE. Los intervalos de confianza proyectados por el modelo incluyen la esperanza de vida proyectada por el INE en un porcentaje superior al 95% en el global. Por años el porcentaje de éxito es especialmente relevante en un horizonte de 2 años, para 2010, donde se sitúa de acuerdo con el ejercicio el tiempo presente y que supera el 99%. Este porcentaje de éxito decrece con el tiempo aunque hasta los diez años esta por encima del 90%. Por provincias el éxito supera el 95% salvo en casos excepcionales sin motivo aparente. La distribución provincial del éxito en este ejercicio sobre la coincidencia con la proyección es aparentemente independiente de la distribución del ejercicio de la validación.

Resumiendo, tanto la validación con datos observados como la comparación con la proyección proporciona porcentajes de éxito notables, un poco mejores en el caso de la comparación.

En conclusión, la validación del modelo propuesto en el apartado 4.5 obtiene unos resultados satisfactorios respecto a los valores observados y comparables con la proyección a corto plazo proporcionada por el INE. Además facilita un intervalo de confianza para la estimación de la esperanza de vida provincial. Este intervalo de confianza constituye en esencia una medida de la bondad de la predicción de la que carecen las proyecciones recogidas en la bibliografía hasta la fecha.

Capítulo 6

Conclusiones y líneas abiertas de investigación

- 6.1 Introducción
- 6.2 Difusión de la investigación
- 6.3 Conclusiones
- 6.4 Líneas abiertas de investigación

6.1. Introducción

El objeto de esta Tesis es un detallado estudio de la incidencia de la mortalidad, con especial atención a las poblaciones de pequeño tamaño. El trabajo se concentró en los cuatro elementos capitales: la medición y el análisis de la mortalidad, la especificación del modelo y su validación. Con este fin, se fijó un plan estratégico con un camino que pasa por generar un marco científico de garantías para la proyección de la mortalidad. Ahora, la Tesis alcanza sus últimos objetivos con la presentación de las conclusiones y la exposición de las líneas abiertas de investigación.

Los resultados de la Tesis son útiles para cualquier subpoblación. El estudio se aplica en el caso particular de las provincias de España para las que se derivan una serie de resultados prácticos. Todos estos resultados son extensibles a otras poblaciones y está especialmente diseñada para aquellas poblaciones que adolezcan de falta de consistencia por carecer de tamaño suficiente, por estar expuestas a fluctuaciones bruscas o cualquier otro motivo. Es suficiente con que sean parte integrante de una población mayor que funcionará como población de referencia y respecto de la cual se calcularán los diferenciales que dan lugar a los intervalos de confianza.

Este capítulo describe la trayectoria de mis investigaciones hasta la fecha y recopila las conclusiones y las líneas de investigación abiertas en las distintas etapas que se han ido superando a lo largo del trabajo y que se exponen en los sucesivos capítulos del mismo.

6.2 Difusión de la investigación

Esta Tesis es la culminación de un grupo de trabajos que se han desarrollado a lo largo de los últimos seis años desde que entré a formar parte del Instituto Nacional de Estadística, en 2008, como Estadístico Facultativo en el Área de Análisis y Proyecciones Demográficas. Desde entonces me vengo ocupando de la medición,

análisis, estimación y proyección de la mortalidad tanto a nivel nacional como en la desagregación por Comunidades Autónomas y Provincias.

El primer proyecto en el que participé fue un borrador inicial para la actualización de la metodología del cálculo de las tablas de mortalidad del INE. Utilizando como base esta metodología, en septiembre de 2009 se celebró un seminario para la revisión metodológica del cálculo de las tablas de mortalidad del INE. El objetivo principal del seminario era decidir la opción metodológica de cálculo más adecuada para la producción de tablas de mortalidad tanto a nivel nacional como por Comunidades Autónomas y Provincias. Para ello se elaboró un documento con distintas alternativas y un análisis comparativo descriptivo:

López-Carrión, F.; Cantalapiedra, M.; Muriel, S. (2009): «Tablas de mortalidad de la población de España». Seminario de expertos en relación con la revisión metodológica del cálculo de tablas de mortalidad en el INE. Madrid. 21-22 de septiembre de 2009.

Al año siguiente, el INE presentó en Lisboa el siguiente working paper, con el que se difundieron las conclusiones sobre la mejor alternativa de medición de la incidencia de la mortalidad en España y sus provincias dependiendo de la opción metodológica:

Muriel, S.; Cantalapiedra, M.; López-Carrión, F. (2010): «Towards advanced methods for computing life tables». Work session on Demographic Projections. Lisbon. April 28-30 2010. Instituto Nacional de Estadística. Documentos de trabajo. [http://www.ine.es/ss/Satellite?L=es_ES&c=INEDocTrabajo_C&cid=1259925434773&p=1254735116586&pagename=ProductosYServicios%2FPYSLayout]

En noviembre de 2011 presenté en la Universidad de Alcalá mi Trabajo de Investigación Tutelado, dirigido por José Luís Gutiérrez de Mesa y Luis Felipe Rivera Galicia, génesis de esta Tesis:

López-Carrión, F. (2011): «Revisión Crítica de la medición de la mortalidad en España. Distintas aproximaciones a las tablas de mortalidad». Trabajo de Investigación Tutelado. Universidad de Alcalá de Henares.

Aprovechando algunos de los primeros resultados de esta Tesis, presenté en la *European Population Conference* (Conferencia Europea de Población), celebrada en Estocolmo en junio de 2012, un trabajo que se aceptó para la sesión 48, Measures of mortality, presidida por Dalkhat M. Ediev¹:

López-Carrión, F.; Gutierrez de Mesa, J.L. y Rivera, L.F. (2012): «Mortality confidence intervals: a measure of prediction goodness». 2012 European Population Conference. Stockholm. June 13-16 2012. [<http://epc2012.princeton.edu/papers/120267>]

En noviembre de 2013, Luis Felipe Rivera Galicia presentó en la XXXIX Reunión de Estudios Regionales, celebrada en Oviedo, un avance de algunas de las conclusiones del apartado 3.4 de esta Tesis sobre la clasificación de las provincias de España por su esperanza de vida:

Rivera, L.F.; López, F. y Gutiérrez de Mesa, J.L. (2013): « Análisis de la Esperanza de Vida en las Provincias Españolas. 1991-2010». XXXIX Reunión de Estudios Regionales. Oviedo, España. 21-22 de noviembre de 2013. [<http://www.reunionesdeestudiosregionales.org/Oviedo2013/htdocs/pdf/p827.pdf>]

¹ Senior Scientist at the Vienna Institute of Demography of Austrian Academy of Sciences.

Finalmente, este documento recoge los resultados y conclusiones junto con todo el trabajo realizado a lo largo de estos últimos seis años. La difusión de mi investigación se completa con la conclusión de esta Tesis en abril de 2014 y con la intención de publicar en revistas especializadas los resultados más interesantes de la misma.

6.3 Conclusiones

Esta Tesis ha demostrado la importancia de la dimensión geográfica en la incidencia de la mortalidad. Se han propuesto dos nuevos modelos para la proyección de la mortalidad provincial. Un modelo para el riesgo de muerte que salva el problema del aumento de la frecuencia observada de ceros y otro modelo que proporciona, además de una estimación puntual, una estimación por intervalos eficaz, eficiente y útil.

Ambos modelos se pueden aplicar a cualquier subpoblación incluida en una población de referencia con la única condición de que la forma de medir la incidencia de la mortalidad sea la misma.

El resultado más destacado de la Tesis es un novedoso protocolo de actuación para la proyección de la mortalidad provincial basado en una metodología relacional encaminado a proporcionar, más allá de la clásica estimación puntual de la mortalidad, una estimación por intervalos. El fruto directo es un intervalo de confianza que constituye en esencia una medida de la bondad de la predicción.

La Tesis sigue un plan estratégico concentrado en los cuatro elementos capitales: la medición y el análisis de la mortalidad, la especificación del modelo y su validación.

El estudio sobre la medición de la mortalidad se trata en el capítulo 2 y da lugar a una serie de importantes conclusiones:

- En el principio, se parte de un estudio descriptivo del fenómeno demográfico de la mortalidad cuya principal conclusión es una alternativa metodológica que ofrece una descripción cuantitativa coherente, sistemática y lo más

completa posible para la medición de la mortalidad aplicable en España, que saca el máximo provecho de las fuentes de información disponible.

- Una conclusión de especial relevancia es la constatación de que la medición de la incidencia de la mortalidad va más allá de un simple ejercicio contable para convertirse en un verdadero ejercicio estadístico de estimación.
- La medida de la mortalidad como ejercicio estadístico parte del interés particular de las personas en conocer su propia esperanza de vida. Por ejemplo, un varón que reside en un núcleo urbano con un alto nivel de contaminación y con un nivel educativo y socio-económico bajo. Un individuo con unas características biológicas, ambientales y sanitarias y unas circunstancias educativas y socio-económicas particulares en un determinado lugar. Los cálculos con una población así definida de un elemento muestral da lugar a una tabla de mortalidad degenerada.
- La solución es ampliar la muestra hasta, por ejemplo, la población de España. Una muestra de un tamaño de más de 46 millones de elementos muestrales que proporciona unas propiedades deseables en términos de consistencia en la estimación de los parámetros.
- Ocurre que en ese camino hacia la consistencia, la muestra pierde representatividad frente a las características y circunstancias del individuo original objeto de interés. La provincia es una etapa intermedia.
- La muestra es más representativa porque las características y circunstancias de sus elementos muestrales son más próximos a las características y circunstancias del individuo original objeto de interés. Lamentablemente, al mismo tiempo que se reduce el tamaño de la muestra, la estimación pierde consistencia y fiabilidad. Nos encontramos con dos extremos. Un extremo proporciona máxima consistencia con el tamaño poblacional más grande y el otro extremo máxima representatividad con el tamaño poblacional más pequeño. Entre estos extremos nos encontramos el caso intermedio de la provincia.

- Otro interesante resultado es una detallada revisión histórica del fenómeno demográfico de la mortalidad a nivel internacional, y particularmente en España. Se concluye la especial dependencia de las tablas de mortalidad respecto de sus fuentes de información. La estimación de la incidencia de la mortalidad se concreta en el cálculo de una tabla de mortalidad como principal instrumento para su medición. La fiabilidad de la medición de la incidencia de la mortalidad depende, de forma crítica, de la calidad de las fuentes de información sobre la población pero sobre todo de las defunciones para las que la estimación es especialmente sensible. Merecen un especial reconocimiento todos los profesionales de la Estadística y de la Demografía que han trabajado a lo largo de la Historia en el estudio de la incidencia de la mortalidad, en muchas ocasiones a pesar de las limitaciones técnicas y las circunstancias dramáticas que han acaecido en España, especialmente a lo largo del siglo XX.
- Otra importante conclusión es la existencia de diferencias clave entre las propiedades de las funciones biométricas de la tabla de mortalidad. Estas diferencias hacen a las funciones que se construyen recursivamente más atractivas por su consistencia en la estimación.
- Se concluye, además, la preferencia de la esperanza de vida, los supervivientes o la población estacionaria, como poblaciones que se construyen recursivamente, frente al riesgo de muerte, la tasa de mortalidad o las defunciones teóricas. Sobre todo en poblaciones de pequeño tamaño para las que el riesgo de muerte, y el resto de funciones independientes en edad, son especialmente sensibles a factores coyunturales, discrecionales, circunstanciales o aleatorios.

Una vez que se cierra el estudio sobre la medición de la mortalidad en el capítulo segundo se pasa a tratar en el capítulo 3 el siguiente elemento del plan estratégico, como es el análisis de la mortalidad en España y sus provincias.

El análisis de las funciones biométricas de las tablas de mortalidad de España y sus provincias conduce a una serie de importantes conclusiones, que se enumeran a continuación:

- En primer lugar, se confirma la importancia del tamaño de la población como estabilizador de la estimación de los parámetros. Se concluye que todas las funciones biométricas tienen un importante componente de contigüidad respecto al tiempo y la edad en las poblaciones de tamaño grande.
- El tiempo de vida medio vivido el último año de vida presenta patrones por generación. Se descarta cualquier dependencia con la población por ser una función que se calcula a partir del *Movimiento Natural de Población*, fuente de las defunciones. Se plantea como posible causa la imputación sistemática de registros incompletos.
- Por otra parte, las funciones biométricas de las provincias presentan discontigüidad tanto en la variable edad como en la variable tiempo, sobre todo en las funciones independientes por edad, como el riesgo, frente a las funciones recursivas, como la esperanza de vida.
- La observación de las tablas y gráficos elaborados lleva a la importante conclusión de la preferencia de la esperanza de vida frente al riesgo como función más estable.
- Una conclusión importante es el alto grado de contigüidad en la esperanza de vida entre las provincias que se traduce en una sensación de continuidad en la superficie determinada por las provincias.
- Además, se establece como conclusión relevante que la distribución geográfica provincial para el riesgo de muerte también presenta contigüidad, aunque en menor grado que en el caso de la esperanza de vida.
- Otra destacada conclusión para la esperanza de vida al nacimiento es la estabilidad de la distribución geográfica provincial relativa en el tiempo. Las provincias del sur y costeras presentan una mortalidad relativa peor que las del norte e interiores. A lo largo del período considerado, se aprecia que se

produce un paulatino desplazamiento de la región con peor mortalidad relativa por la costa hacia el este.

- El estudio descriptivo de los datos concluye con la existencia de una tendencia común de todas las provincias que se traduce en un crecimiento monótono, paulatino y constante de la esperanza de vida del total nacional.
- Otra interesante conclusión es la constatación de que la distribución geográfica relativa es independiente del sexo en el sentido de que la ley que gobierna la mortalidad afecta por igual a las dos poblaciones sin distinción significativa ni discriminación por razones de sexo.
- El estudio y clasificación de las provincias por esperanza de vida para cualquier edad con un análisis de componentes principales tiene como principal conclusión la preponderancia de la esperanza de vida al nacimiento. Este índice explica prácticamente en su totalidad la incidencia de la mortalidad en la población. Se concluye por tanto que la aportación de información para explicar la ley de la mortalidad de las esperanzas de vida a otras edades es marginal.
- Además se concluye que la clasificación de las provincias confirma la relevancia de la distribución geográfica como factor de influencia en la mortalidad.

Con esto se cierra el análisis descriptivo de la mortalidad en España y sus provincias del capítulo tercero.

El siguiente elemento del plan estratégico seguido en la tesis es la especificación de dos novedosos modelos para describir el comportamiento de la incidencia de la mortalidad en las provincias. Este elemento lleva a una serie de importantes conclusiones que se detallan en el capítulo 4:

- La primera conclusión es que la modelización de la esperanza de vida proporciona mejores resultados que la del riesgo de muerte. El riesgo de muerte tiene importantes inconvenientes por estar más expuesto a factores aleatorios o circunstanciales que dificultan su estimación, en concreto, el

aumento de la frecuencia de ceros observados para el riesgo de muerte en las provincias.

- Se presenta la especificación de un modelo que explica el riesgo de muerte provincial con una particular atención a las especiales características detectadas.
- Se concluye con la constatación de las dificultades en el ajuste que plantean las especiales características del riesgo de muerte en poblaciones de pequeño tamaño.
- Partiendo de una estimación puntual se presenta la innovadora alternativa de una estimación por intervalos de confianza para la esperanza de vida provincial. La construcción de este intervalo de confianza constituye en esencia una medida de la bondad de la predicción y una solución fiable, razonable y pragmática a la par que elegante para la problemática de la estimación en pequeñas poblaciones. Tiene el mérito añadido de ser un planteamiento totalmente innovador del que carecen las proyecciones recogidas en la bibliografía hasta la fecha.
- Los extremos del intervalo de confianza permiten satisfacer la actual demanda de diferentes escenarios para la proyección de la mortalidad, usando la esperanza de vida, pudiendo elegir el extremo superior como escenario más favorable y el extremo inferior como escenario menos favorable.
- Se concluye además que el análisis de residuales del modelo es satisfactorio.
- El producto obtenido es un protocolo metodológico que funciona para cualquier subpoblación de una población, con la única condición de que los criterios de medición para los parámetros de la tabla de mortalidad sean iguales.

La validación del modelo es el último elemento del plan estratégico y se recoge en el capítulo quinto de la tesis, cuyas conclusiones son las siguientes:

- La principal conclusión es que el modelo de proyección propuesto para la esperanza de vida provincial es fiable. Los resultados son satisfactorios. La

validación con los datos observados alcanza porcentajes de éxito elevados, por encima del 95% en la mayoría de los casos, para los intervalos de confianza provinciales.

- La validación contra los datos observados presenta por sexos un porcentaje de éxito ligeramente superior para las mujeres. Por años es mayor en el último año validado. Por provincias el porcentaje de éxito es elevado salvo casos excepcionales. Por edad el porcentaje de éxito se mantiene alto y presenta cierta correlación con las defunciones teóricas.
- Otra importante conclusión es la similitud entre las predicciones de esta tesis y las elaboradas por el INE en las proyecciones de población a corto plazo. El porcentaje de éxito en la coincidencia es aún mayor que en los datos observados y superior en la mayoría de las provincias al 98%.
- La comparación frente a las proyecciones del INE presentan por sexos, de nuevo como en la validación de datos observados, un porcentaje de éxito ligeramente superior para las mujeres. Por años el porcentaje de éxito va desde un éxito casi total los primeros años y disminuye lentamente hasta un razonable porcentaje superior al 90% en el horizonte de los 10 años. Por provincias el porcentaje de éxito es elevado salvo casos excepcionales. Por edad el porcentaje de éxito se mantiene alto y presenta cierta correlación con las defunciones teóricas.
- Finalmente, pero no menos importante, se trata de un modelo útil. Un intervalo de confianza demasiado grande serviría de poco y el modelo proporciona intervalos de confianza razonables.

En resumen, se ha alcanzado el objetivo estratégico de esta tesis con un estudio integral de la mortalidad, con especial atención a la mortalidad en poblaciones de pequeño tamaño. Se concluye la importancia de la dimensión geográfica en la incidencia de la mortalidad y se presenta un novedoso modelo de proyección de la esperanza de vida provincial, basado en una estimación por intervalos, cuyo método se puede aplicar a cualquier subpoblación de forma eficaz, eficiente y útil.

6.4. Líneas abiertas de investigación

Este último apartado es una invitación a todas las personas que, con interés e inquietud por aprender, quieran acompañarme en futuras investigaciones sobre la incidencia de la mortalidad.

Al mismo tiempo que avanzaba en mi camino de esta Tesis iba descubriendo el fascinante paisaje de nuevos entornos sin explorar aún, al menos por mí.

A continuación expongo las líneas abiertas de investigación detectadas en esta Tesis con el vivo interés de acometerlas en un futuro cercano y animar a otros investigadores a que me acompañen en la tarea.

- La calidad de la medición de la incidencia de la mortalidad está directamente condicionada a la calidad de las fuentes de información disponibles. Una primera línea de investigación abierta consiste en suplir las deficiencias del cálculo de la exposición al riesgo de la población en España. En concreto, se trata de generar un sistema de información geográfica que integre los datos del padrón y que funcione a modo de censo continuo. El INE está impulsando ya esta línea de trabajo a través del proyecto *Sistema Integrado de Gestión de Población y Territorio*, dirigido por la Jefa de Área Nazaret García Seijas, desde la subdirección de Tecnologías de la Comunicación y la Información. Esta línea permitirá calcular el valor exacto de la exposición al riesgo de cada individuo de la población objeto de estudio. El futuro apunta claramente en la dirección de esta línea de investigación con algunos prometedores resultados ya obtenidos en algunos países e incluso Comunidades Autónomas en España, en este caso contruidos en base al padrón municipal.
- Las tablas de mortalidad dependen de la función biométrica que se utilice para reconstruirla. La metodología HMD parte de la tasa de mortalidad para calcular el resto de funciones. También podemos partir de $l(x)$, de $e(x)$ o de cualquier otra función de la tabla de mortalidad. La cuestión particularmente interesante es estudiar la robustez de la estimación del resto de las funciones. Necesitamos que la tabla de mortalidad sea un estimador robusto. Es decir,

que todos los indicadores miden la misma intensidad de la mortalidad. Confirmar que una ligera desviación en la estimación de una función biométrica no nos lleve a que el resto de las funciones de la tabla de mortalidad cambien abruptamente.

- En cuanto al estudio de las distintas funciones biométricas de la tabla de mortalidad, queda comparar la esperanza de vida frente a la serie de supervivientes o la población estacionaria. Por construcción la esperanza de vida tiene un comportamiento más estable según se baja en edad y los supervivientes según se sube pero nos interesa saber más. Es necesario realizar un estudio comparativo entre todas las funciones biométricas de la tabla de mortalidad.
- Durante toda la tesis se ha hecho referencia al tamaño de la población según fuera grande o pequeño. Más grande significa una estimación más consistente y estable. Más pequeño implica una estimación más representativa. La virtud se encuentra en el punto medio que satisface ambas aspiraciones. Se trata de definir pequeño cuando hablamos del tamaño de una población, es decir, cuál es el tamaño mínimo para que tenga sentido el cálculo de la tabla de mortalidad de una población. La construcción de la muestra a partir de la que se estima la incidencia de la mortalidad es un ejercicio de inferencia estadística que implica un estudio por edades. Una edad con un $q(x)$ inferior a la diez milésima necesita un tamaño de muestra mayor. Para otras edades con otros valores de $q(x)$ mayores el tamaño de la muestra puede ser menor. El tamaño de la muestra es diferente para cada edad y el tamaño de la población es el agregado de todas las muestras para todas las edades.
- Se descubre en este punto la esencia que se esconde tras el cálculo de la esperanza de vida. Nos encontramos ante el conflicto entre consistencia y representatividad de la estimación. Cuanto más grande sea la población utilizada, mayor será la consistencia de la estimación, pero esto lleva a diluir la identidad del individuo objeto de estudio, el hecho diferencial particular. Lo

que interesa no es tanto un indicador medio de la esperanza de vida en una población grande y, por tanto, más heterogénea e inevitablemente imprecisa. El interés está en realidad en la medida de la mortalidad del individuo real o teórico objeto de estudio en un lugar puntual concreto bajo unas características particulares singulares y unas circunstancias determinadas. Un estudio que incluye, además, la definición de una distancia para la selección de los elementos más próximos al individuo de referencia objeto de estudio. Una distancia que debe incluir un componente euclídeo de cercanía pero también deberá valorar otras variables como el tipo de medio, rural o urbano, el nivel educativo, además del sexo, variables que midan la exposición a factores de riesgo, etc., dependiendo del interés del estudio.

- Como corolario de esta línea de estudio se plantea otra línea de investigación, que retoma la idea inducida por la continuidad del mapa de esperanzas de vida provincial observado, de elaboración de un mapa de relieve con la esperanza de vida.
- La extensión de la contigüidad geográfica observada en la esperanza de vida induce a estudiar la continuidad. Es necesario incorporar la georreferenciación de los individuos para el cálculo de la esperanza de vida, que pasa de ser una variable asociada a un territorio a estar asociada a un punto geográfico. El objetivo es construir un mapa de superficie continuo como un relieve montañoso para la esperanza de vida, el riesgo de muerte, los supervivientes y el resto de las funciones biométricas.
- Otra línea de investigación abierta es replicar el estudio realizado con otras subpoblaciones que constituyan al agregarse una población de referencia. Para empezar, la población de Iberia, como agregación de España y Portugal con una desagregación territorial de acuerdo con sus respectivas divisiones administrativas, en España las provincias y en Portugal, los distritos o incluso las provincias definidas en la Constitución Portuguesa de 1936. Se puede replicar el estudio con los países de la Unión Europea frente al agregado de la Unión Europea. Otra posibilidad son los Estados Unidos de América y sus

estados. Sería interesante estudiar el comportamiento del modelo en poblaciones cuyas fuentes de información sean poco fiables a nivel regional y para las que se disponga de tablas tipo de garantía que sirvan como población de referencia.

- La creciente demanda de información sobre longevos, superlongevos, centenarios, semisupercentenarios y supercentenarios presiona a las Entidades Estadísticas a prolongar la edad de cierre de las tablas de mortalidad. Es necesario determinar una edad de cierre de la tabla de mortalidad y obtener una estimación de la esperanza de vida a la edad de cierre.
- El análisis descriptivo de la función de años vividos el último año de vida, $a(x)$, induce a pensar que el Registro de Defunciones del Movimiento Natural de Población presenta errores sistemáticos en la imputación. Sería necesario realizar un estudio de estos patrones observados por generación. Una primera aproximación al problema es aislar una generación que presente este patrón y estudiar la distribución de $a(x)$, por fecha de nacimiento y fecha de defunción, estudiando si existe correlación entre estas variables, lo que llevaría a sospechar sobre criterios sistemáticos de imputación por falta de respuesta u otros motivos.
- Para la Demografía Histórica y la Sociología queda la explicación de la evolución a lo largo del tiempo de las distintas funciones biométricas. La incidencia de la mortalidad nacional depende de la edad y del tiempo. Es necesario estudiar la evolución de la mortalidad dependiendo de estos factores. Respecto de la mortalidad por provincias, es necesario dar respuesta a las diferencias y evolución geográfica relativa que se ha observado. También se han constatado las diferencias entre sexos. El análisis de la sobremortalidad y las diferencias en mortalidad que, aunque se reducen aparentemente se mantiene en el tiempo. Es necesario contrastar la hipótesis de convergencia entre los sexos o lo que es lo mismo la igualdad entre sexos frente a la incidencia de la mortalidad. Al mismo tiempo se debería contrastar

la hipótesis de convergencia entre las distintas provincias. Esta Tesis no encuentra evidencia suficiente para aceptar la convergencia de la mortalidad de las provincias en el tiempo. Adicionalmente, existen otros posibles contrastes, discriminando por estado civil, nivel educativo, situación social o estatus económico.

- En la clasificación de las provincias por el análisis combinado se optó por un proceso en dos fases: primero discriminar respecto al total nacional y luego en una segunda etapa con las provincias ambiguas proceder a una clasificación por análisis de conglomerados porque empíricamente daba mejores resultados que una clasificación por conglomerados directa. Es necesario comparar estos dos métodos de clasificación y aplicar otros métodos de clasificación del Análisis de Datos o de la minería de datos.
- Es interesante estudiar otros modelos alternativos para la proyección del riesgo de muerte que aporten otras soluciones al problema del aumento de la frecuencia de ceros. Una solución que proporcione un intervalo de confianza para la proyección e incluya al cero en el intervalo de confianza como punto discreto asignándole una probabilidad.
- En el futuro, será posible ampliar el período de validación del modelo para la esperanza de vida provincial. La validación contra datos de 3 años es una primera aproximación. Es necesario probar con más años, a medida que se disponga de más datos. Actualmente, sólo se puede incorporar el año 2012.
- Se ha detectado que la amplitud del intervalo de confianza para la esperanza de vida provincial está relacionado con el tamaño poblacional de la provincia, así como con el valor de la esperanza de vida nacional. El siguiente paso es realizar un análisis descriptivo y especificar un modelo que explique la amplitud del intervalo en función del tamaño de la población y de la esperanza de vida nacional.
- Existe una amplia bibliografía sobre modelos que se pueden utilizar para explicar el comportamiento y evolución en el tiempo de la mortalidad, tradicionalmente la tasa de mortalidad o el riesgo de muerte. En este sentido,

puede ser interesante realizar un estudio comparativo entre las diversas alternativas (modelos ARIMA, el modelo Lee-Carter, modelos de P-splines, modelos de B-splines, etc.) aplicadas a la modelización de las proyecciones de mortalidad.

- Sería interesante ampliar el abanico de posibilidades en cuanto a la proyección de la mortalidad, considerando otras funciones biométricas diferentes a la tasa o el riesgo de mortalidad. Son especialmente sugerentes las funciones biométricas de las defunciones teóricas, los supervivientes, la población estacionaria o la esperanza de vida. Estas funciones podrían dar resultados satisfactorios.

Como se puede apreciar, esta Tesis abre un amplio abanico de nuevos campos por analizar a través de estas líneas abiertas de investigación con el objeto de mejorar la comprensión de la incidencia de la mortalidad en las poblaciones.

Como corolario, es necesario señalar las importantes aplicaciones del modelo propuesto en esta Tesis para la previsión en la consolidación del Sistema de la Seguridad Social española. Su desarrollo constituye un futuro proyecto, ambicioso y actualmente en sus primeras fases de análisis, que estudiará la proyección de la población activa, la población pensionista y la evolución de los índices de sostenibilidad aprovechando las conclusiones y resultados de esta Tesis. El lector interesado en una primera aproximación al tema puede consultar el capítulo 4 de la tesis doctoral de Gutiérrez de Mesa (2002) donde encontrará un completo y riguroso estudio sobre aplicaciones de modelos demográficos de previsión a la consolidación de la Seguridad Social.

Referencias

AHCAN, A.; MEDVED, D.; OLIVIERI, A.; PITACCO, E. (2014): «Forecasting mortality for small populations by mixing mortality data». Insurance: Mathematics and Economics, Volume 54, pp. 12-27.

BRASS, W. (1971): «On the scale of mortality» In W. Brass (ed.). Biological Aspects of Demography. London: Taylor and Francis.

BRASS, W. (1975): «On Methods for estimating fertility and mortality from limited and defective data» Duxbury Press. North Scituate, Mass.

CAMARDA, C. G. (2008): «Goodness of fit in models for mortality data» Tesis Doctoral. Universidad Carlos III de Madrid, Departamento de Estadística.

CANTALAPIEDRA, M. (2001): «Proyecciones», Fuentes Estadísticas, 55, 14.

CANTALAPIEDRA, M. (2002): «El envejecimiento de la población en España», Fuentes Estadísticas, 68, 26-27.

CANUDAS-ROMO, V.; BECKER, S. (2011): «The crossover between life expectancies at birth and at age one: The imbalance in the life table» Demographic Research Volume 24, 2011, Pages 113-144.

CANUDAS-ROMO, V. (2010): «Three measures of longevity: Time trends and record values» Demography 47(2): 299-312.

CANUDAS-ROMO, V.; GLEI, D.; GÓMEZ-REDONDO, R.; COELHO, E.; BOE, C. (2008): «Mortality changes in the Iberian Peninsula in the last decades of the twentieth century» Population (english edition), 2008, vol. 63, issue 2, pages 319-343.

COHEN, J. E. (1977): «Ergodicity of age structure in populations with Markovian vital rates». III. Finite-state moments and growth rate; an illustration, Ah. Appl. Probab. 9. 462-475.

DUPÂQUIER, J. (1996): «L'invention de la table de mortalité, de Graunt à Wargentin, 1662-1766» Collections Sociologies. Paris: Presses Universitaires de France, 1996. 177 pp. Tables.

FRENCH, D. (2014): «International mortality modelling — An economic perspective», Economic Letters, 122(2), pp.182-186.

GABRIEL, W. (1992): «Survival of Small Populations under Demographic Stochasticity» Theoretical population biology: pp. 44-47.

GIROSI, F.; KING G. (2008): «Demographic Forecasting» Princeton, Princeton University Press.

GIROSI, F.; KING G. (2007): «Understanding the Lee-Carter Mortality Forecasting Method» [<http://gking.harvard.edu/files/gking/files/lc.pdf>].

GLEI, D.; GÓMEZ-REDONDO, R.; ARGÜESO, A.; CANUDAS-ROMO, V. (2008): «About mortality data for Spain», Updated: February 19, 2008. Background and Documentation, The Human Mortality Database. (<http://www.mortality.org>).

GOERLICH, F. J. (2008): «Las tablas de mortalidad del Instituto Nacional de Estadística: 1900-1901 a 2004-2005. Recopilación Crítica», Estadística Española, 169, pp. 523-569.

GOERLICH, F. J.; MAS, M. (DIRS.); AZAGRA, J.; CHORÉN, P. (2006): «La Localización de la Población Española sobre el Territorio. Un siglo de cambios. Un estudio basado en series homogéneas» (1900-2001). Madrid: Fundación BBVA.

GÓMEZ REDONDO, R.; GARCÍA GONZALEZ, J.M. (2011): «Longevidad y salud a edades avanzadas», in Gómez Redondo, R. (dir.), Salud, demografía y sociedad en la población anciana, CES, Alianza Editorial, Madrid, pp. 57-84.

GOMPERTZ, B. (1825): «On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies». Philosophical Transactions of the Royal Society of London 115: 513–585. doi:10.1098/rstl.1825.0026.

GRAUNT, J. (1662): «Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality». (Recogido en Journal of the Institute of Actuaries, 1964, 90, 1-61 con una Introducción de B. Benjamin). London.

GUIJARRO, M.; PELÁEZ, O (2008): «La longevidad globalizada: un análisis de la esperanza de vida en España (1900-2050)», Scripta Nova Vol. XII, núm. 260.

GUIJARRO, M.; PELÁEZ, O (2009): «Proyección a largo plazo de la esperanza de vida en España», Estadística Española Vol. 51, núm. 170, 2009, pp. 193-220.

GUTIÉRREZ DE MESA, J.L. (2002): «Revisión crítica de los actuales modelos de proyección de la población». Tesis Doctoral. Universidad de Alcalá. Cap. 4, pp. 237-258.

GUTTERMAN, S.; VANDERHOOF, I. (1998): «Forecasting Changes in Mortality: A Search for a Law of Causes and Effects » North American Actuarial Journal Volume 2, Issue 4.

HALLEY, E. (1693): «An estimate of the degrees of the mortality of mankind», Philosophical Transactions, 17, 596-610, 653-656.

HELIGMAN, L.; POLLARD, J.H. (1980): «The age pattern of mortality» Journal of the Institute of Actuaries, 1980 - Cambridge University Press.

HOEM, J. M. (1977): «A Markov chain model for working life tables» Scandinavian Actuarial Journal. Volume 1977, issue 1.

HUMAN MORTALITY DATABASE (2000): «Tablas de Mortalidad». Berkeley. [<http://www.mortality.org/>].

INSTITUTO NACIONAL DE ESTADÍSTICA (1941): «Curva logística de la población de España». Boletín de Estadística nº 12. Octubre-diciembre. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1946): «Tablas de mortalidad de la población española. Años 1930-31». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1951): «Tablas de mortalidad de la población española. Años 1900 a 1940». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1956): «Curva logística de la población de España». Serie A. Demografía 1. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1958): «Tablas abreviadas de mortalidad de la población española. Año 1950». Serie A. Demografía – 2. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1959): «Evolución de las tasas específicas de mortalidad de la población española. Periodo 1940 a 1950». Serie A. Demografía – 3. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1960a): «Tablas de mortalidad de la población española. Año 1950». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1960b): «Esquemas de supervivencia de la población española. Periodo 1929 a 1957». Serie A. Demografía – 4. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1963): «Tablas abreviadas de mortalidad de la población española. Año 1960». Investigaciones Demográficas y Sociales. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1964): «Esquemas de supervivencia de la población española. Periodo 1957 a 1960». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1976): «Panorámica demográfica. Análisis, estructura y proyecciones de la población española». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1977a): «Tablas de mortalidad de la población española. Años 1960-70». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1977b): «Tablas abreviadas de mortalidad de la población española. Año 1970». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1978): «Tablas de mortalidad provinciales (1969-72). Año 1970». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1979): «Evolución de la población de España. 1961-1978». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1981a): «Tablas de mortalidad de la población española. Años 1975-76». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1981b): «Proyección de la población española para el periodo 1978-1995». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1987): «Proyección de la población española para el periodo 1980-2010. Resultados para el Conjunto Nacional». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1988a): «Tablas de mortalidad de la población española por Comunidades Autónomas. Años 1970-1975-1980». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1988b): «Tablas de mortalidad de la población española. Años 1980-81». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1988c): «Proyección de la población española para el periodo 1980-2010. Resultados por Comunidades Autónomas». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1991): «Tablas de mortalidad de la población española. Años 1985-86». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1993): «Tablas de mortalidad de la población española. Años 1990-91». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1995): «Proyecciones de la población de España calculadas a partir del Censo de Población de 1991». Mortalidad. p. 13-19. Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1997): «Tablas de Mortalidad de la Población Española. Años 1985 y 1990. Resultados por Comunidades Autónomas». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1998): «Tablas de mortalidad de la población española. Años 1994-1995». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1999a): «Tablas de Mortalidad de la Población Española 1994-1995. Resultados por Comunidades Autónomas». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (1999b): «Tablas de Mortalidad de la Población Española 1996-1997». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (2002a): «Tablas de Mortalidad de la Población de España. 1998-1999». Madrid. INE. [<http://www.ine.es>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2002b): «Tablas de Mortalidad de la Población de España. 1998-1999. Resultados por Comunidades Autónomas», Madrid. INE. [<http://www.ine.es>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2003): «Proyecciones de la población de España calculadas a partir del Censo de Población de 2001». Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (2007): «Metodología empleada en el cálculo de las Tablas de Mortalidad de la Población de España. 1992-2005», Madrid. INE. [http://www.ine.es/daco/daco42/mortalidad/metodo_9205.pdf].

INSTITUTO NACIONAL DE ESTADÍSTICA (2008): «Estimaciones de Población Actual. Metodología hasta el 1 de enero de 2009», Madrid. INE. [<http://www.ine.es/daco/daco43/epoba/metodo.pdf>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2009a): «Tablas de mortalidad. Metodología», Madrid. INE. [http://www.ine.es/daco/daco42/mortalidad/metodo_9107.pdf].

INSTITUTO NACIONAL DE ESTADÍSTICA (2009b): «Propuesta para la celebración de un Seminario de expertos en relación con la revisión metodológica del cálculo de tablas de mortalidad en el INE. », Madrid. INE.

INSTITUTO NACIONAL DE ESTADÍSTICA (2010): «Proyecciones de Población de España a Corto Plazo (2010-2020). Metodología», Madrid. INE. [http://www.ine.es/metodologia/t20/t20269_m2010.pdf].

INSTITUTO NACIONAL DE ESTADÍSTICA (2011): «Proyecciones de Población de España a Corto Plazo (2011-2021). Metodología», Madrid. INE. [http://www.ine.es/metodologia/t20/t20269_m2011.pdf].

INSTITUTO NACIONAL DE ESTADÍSTICA (2012a): «Movimiento natural de la población. Datos provisionales. Año 2012. Defunciones por edad, mes y sexo», Madrid. INE. [<http://www.ine.es>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2012b): «Estimaciones de Población Actual. Metodología Octubre de 2012», Madrid. INE. [<http://www.ine.es/metodologia/t20/t2030259.pdf>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2012c): «Proyecciones de Población de España a Corto Plazo (2012-2022). Metodología», Madrid. INE. [http://www.ine.es/metodologia/t20/t20269_m2012.pdf].

INSTITUTO NACIONAL DE ESTADÍSTICA (2013a): «Tablas de mortalidad. Serie revisada 2002-2011 (18 de junio de 2013)», Madrid. INE. [<http://www.ine.es>].

INSTITUTO NACIONAL DE ESTADÍSTICA (2013b): «Proyecciones de Población de España a Corto Plazo (2013-2023). Metodología», Madrid. INE. [http://www.ine.es/metodologia/t20/t20269_m2013.pdf].

JEUNE, B.; VAUPEL, J.W. -Eds.- (1999): «Validation of exceptional longevity Odense Monographs on Population Aging» No. 6. (249 pgs).

JONES, O. R.; SCHEUERLEIN, A.; SALGUERO-GOMEZ, R. ; CAMARDA, C. G.; SCHAIBLE, R.; CASPER, B. B.; DAHLGREN, J. P.; EHRLIN, J.; GARCÍA, M. B.; MENGES, E. S.; QUINTANA-ASCENCIO, P. F.; CASWELL, H.; BAUDISCH, A. ; VAUPEL, J. W. (2014): Diversity of ageing across the tree of life. *Nature*. [<http://www.nature.com/nature/journal/v505/n7482/full/nature12789.html>].

KANAROGLOU, P. S.; MAOH, H.; NEWBOLD, B.; SCOTT, D.; PÁEZ, A. (2009): «A demographic model for small area population projections: an application to the Census Metropolitan Area of Hamilton in Ontario, Canada» *Environment and Planning: volume 41(4)* pages 964–979.

KEYFITZ, N. (1982): «Choice of function for mortality analysis: Effective forecasting depends on a minimum parameter representation». *Theoretical population biology*. Volume 21, Issue 3, pp. 329–352.

KEYFITZ, N. (1979): «Introducción a las matemáticas de población». Santiago de Chile. Centro Latinoamericano de Demografía (CELADE) (original en versión inglesa de 1968, «Introduction to the Mathematics of Population», Addison-Wesley Publishing Company, Inc. New York).

KEYFITZ, N. (1970): «Finding probabilities from observed rates or how to make a life table». *The American Statistician*, 24, 1, (February), 28-33.

KEYFITZ, N. (1968): «A life table that agrees with the data II». *Journal of the American Statistical Association*, 63, 324, 1252-1268.

KEYFITZ, N. (1966): «A life table that agrees with the data». *Journal of the American Statistical Association*, 61, 314, 305-312.

KRZANOWSKI, W.J. (1979): «Between groups comparison of Principal Components» *Journal of the American Statistical Association*, Vol. 74, pp. 703-707.

LEE, R. D.; CARTER, L. (1992): «Modeling and forecasting the time series of US mortality», en J. Am. Stat. Assoc. , 87, pp. 659-671.

LEGUINA, J. (1989): «Fundamentos de demografía». Madrid. Siglo XXI (4ª ed.).

LEXIS, W. (1875): «Einleitung ein die Theorie der Bevölkerungs-Statistik». Strasbourg, Trubner.

LÓPEZ-CARRIÓN, F; GUTIERREZ DE MESA, J.L.; RIVERA, L.F. (2012): «Mortality Confidence Intervals: A measure of prediction goodness». Princeton.
[<http://epc2012.princeton.edu/papers/120267>]

MALTHUS, T.R. (1798): «An Essay on the Principle of Population». London: J. Johnson, in St. Paul's Church-yard.

MYERS, R. (1990): «Classical and Modern Regression With Applications» Duxbury classic series.

MILNE, J. (1815): «A Treatise on the Valuation of Annuities and Assurances on Lives and Survivorships; on the Construction of Tables of Mortality; and on the Probabilities and Expectations of Life», London, 1815, 2 vols.

PRESTON, S. H.; HEUVELINE, P.; GUILLOT, M. (2001): «Demography. Measuring and Modelling Population Processes». Oxford. Blackwell.

POLLARD, J. H. (1988): «On the decomposition of changes in expectation of life and differentials in life expectancy», Demography, 25, 2, 265-276.

POLLARD, J. H. (1982) «The expectation of life and its relationship to mortality», Journal of the Institute of Actuaries, 109, 225-240.

RENCHE, A (1995): «Methods of Multivariate Analysis», Wiley Series in Probability and Mathematical Statistics.

RIVERA, L.; LÓPEZ-CARRIÓN, F.; GUTIERREZ DE MESA, J.L. (2013): «Análisis de la Esperanza de Vida en las Provincias Españolas. 1991-2010». XXXIX Reunión de Estudios Regionales. Oviedo, España. 21-22 de noviembre de 2013.
[<http://www.reunionesdeestudiosregionales.org/Oviedo2013/htdocs/pdf/p827.pdf>]

TABEAU, E.; BERG, A; HEATHCOTE, H. (2001): «Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective». Kluwer

THATCHER, A. R.; KANNISTO, V.; VAUPEL, J. W. (1998): «The Force of Mortality at Ages 80 to 120». Odense, Denmark. Odense University Press.

VICIANA, F. (2010): «Base de Datos Longitudinal de Población de Andalucía (BDLPA): Modelo de datos y sistema de gestión», Jornadas de Estadística de las Comunidades. [<http://www.jecas.org/ponencias/jueves/tarde/desarrollosII/BDlogitudinalIAE.pdf>]

WARGENTIN, P (1766): «Tables of mortality based upon the Swedish population». Reprint of an article published in the "Kongl. vetenskaps academiens handlingar fêor år 1766" with reproduction of original t.-p., followed by English translation with special t.-p English t.-p. and cover-title read: Mortality in Sweden, according to the "Tabellverket" (General register office) "A short notice concerning Wargentin's position in mortality investigation and an abbreviated biography": p. 63-68.

WILMOTH, J.R.; ZUREICK, S.; CANUDAS-ROMO, V.; INOUE, M.; SAWYER, C. (2012): «A Flexible Two-Dimensional Mortality Model for Use in Indirect Estimation» Population Studies, 66(1), pp. 1-28.

WILMOTH, J. R.; ANDREEV, K.; JDANOV, D; GLEI, D.A. (2007): «Methods protocol for the Human Mortality Database», mimeo (May 31), Version 5. [<http://www.mortality.org>]. (Versiones anteriores: February 1, 2005 (Version 4); November 29, 2004 (Version 3); October 1, 2002).

WRYCZA, T.F.; BAUDISCH, A. (2012): «How life expectancy varies with perturbations in age-specific mortality» Demografic Research. Vol 27, Artículo 13, Páginas 365-376.