



Universidad de Granada

DEPARTAMENTO DE ESTADÍSTICA E
INVESTIGACIÓN OPERATIVA

**APLICACIÓN DE UN MODELO EDAD-
PERIODO- COHORTE A LOS SUICIDIOS DE
INGLATERRA Y GALES**

Alumno: Christian Hernández Guedes

Tutora: María Dolores Martínez Miranda

Septiembre 2021

Índice general

1. Introducción	9
2. Motivación	11
2.1. El problema de los suicidios en Inglaterra y Gales	12
2.1.1. Definición de suicidio	14
2.1.2. Patrones de suicidio en función de la variable edad	15
2.1.3. Otras características de los suicidios	17
2.2. Revisión de la literatura	17
2.3. Descripción de la base de datos	24
3. Fundamentos teóricos de los modelos edad- periodo- cohorte	29
3.1. Formulación teórica del modelo	29
3.2. El problema de identificación	31
3.3. Procedimientos para abordar el problema de identificación	34
3.3.1. Parametrización canónica	35
3.3.1.1. Interpretación de los efectos temporales	37
3.3.2. Identificación del modelo mediante la imposición de restricciones	38
3.3.2.1. Imposición de restricciones sobre los niveles	38
3.3.2.2. Imposición de restricciones sobre las pendientes	39
3.3.2.3. Geometría de los modelos edad- periodo- cohorte	42
3.3.3. El papel de las funciones estimables	44
3.3.3.1. Aproximación $l'sv$ y obtención de funciones estimables	45
3.3.4. Métodos de identificación bayesianos	49
3.3.5. Conclusiones respecto al problema de identificación	50
3.4. Estimación del modelo	50
3.5. Descomposición de la varianza y extensiones	51
3.5.1. Análisis edad- periodo- cohorte de la varianza (aproximación ANOVA)	51
3.5.2. Submodelos	52
3.5.2.1. Modelos edad- cohorte	52
3.5.2.2. Submodelos de tipo lineal	52
3.5.2.3. Submodelos surgidos a partir de distintas formas funcionales	53
3.5.3. Extensiones de los modelos edad- periodo- cohorte	54
3.5.3.1. Modelos edad- periodo- cohorte mixtos	54
3.5.3.2. Modelos edad- periodo- cohorte jerárquicos	55
3.6. Cuestiones que pueden, o no pueden, resolverse	56
3.7. Alternativas a los modelos edad- periodo- cohorte	57
3.7.1. Análisis gráfico	57
3.7.2. Aproximación factor- característica	57
3.7.2.1. Aproximación por características usando un factor	58
3.7.2.2. Aproximación por características usando dos o más factores	59
3.7.2.3. Comparación de la descomposición de la varianza	59

4. Paquete apc	61
4.1. Creación del objeto de datos	62
4.2. Exploración de los datos	65
4.3. Estimación del modelo	75
4.4. Inferencia sobre el modelo	83
5. Aplicación de un <i>APC model</i> a las cifras y las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019)	85
5.1. Aplicación de un <i>APC model</i> a las cifras de suicidios totales	85
5.1.1. Creación del objeto de datos	85
5.1.2. Exploración de los datos	87
5.1.3. Estimación del modelo	92
5.1.4. Inferencia sobre el modelo	96
5.2. Aplicación de un <i>APC model</i> a las tasas de mortalidad por suicidio	97
5.2.1. Creación del objeto de datos	98
5.2.2. Exploración de los datos	99
5.2.3. Estimación del modelo	102
5.2.4. Inferencia sobre el modelo	105
5.3. Aplicación de un <i>AC model</i> a las tasas de mortalidad por suicidio	107
5.4. Estudio comparativo de las tasas de mortalidad por suicidio de hombres y mujeres	109
5.4.1. Creación de los objetos de datos	109
5.4.2. Exploración de los datos	112
5.4.3. Estimación de los modelos	116
5.4.4. Inferencia sobre el modelo	121
5.5. Discusión de los resultados	123
6. Conclusiones y limitaciones del estudio	127
Bibliografía	129
A. Código en R	137
A.1. Tratamiento de la base de datos original	137
A.1.1. Cifras de suicidios totales	137
A.1.2. Tasas de mortalidad por suicidio	140
A.2. Aplicación de los modelos edad- periodo- cohorte mediante el paquete apc	142
A.2.1. Cifras de suicidios totales	142
A.2.2. Tasas de mortalidad por suicidio	143
A.2.3. Tasas de mortalidad por suicidio (segmentando por sexos)	143

Índice de figuras

2.1.	Ratios de mortalidad por suicidio, registradas en Inglaterra y Gales, entre 1981 y 2019. Se considera, únicamente, a la población mayor de 10 años. Fuente: <i>Office for National Statistics</i> .	12
2.2.	Ratios de mortalidad por suicidio registradas en Inglaterra entre 1981 y 2019; considerando, únicamente, a la población mayor de 10 años. Fuente: <i>Office for National Statistics</i> .	13
2.3.	Ratios de muertes por suicidio, registradas en Gales, entre 1981 y 2019. Fuente: <i>Office for National Statistics</i> .	13
2.4.	Ratios de mortalidad por suicidios registradas en 2019; considerando, únicamente, los fallecimientos de varones. Fuente: <i>Office for National Statistics</i> .	14
2.5.	Ratios de mortalidad por suicidio registradas en 2019; considerando, únicamente, los fallecimientos de mujeres. Fuente: <i>Office for National Statistics</i> .	14
2.6.	Ratios de mortalidad por suicidio considerando la variable edad agrupada en intervalos de 5 años. Datos recolectados en Inglaterra y Galés para el año 2019. Fuente: <i>Office for National Statistics</i> .	16
2.7.	Evolución de las tasas de mortalidad por suicidio en el segmento de varones, considerando las edades agrupadas en distintas franjas. Fuente: <i>Office for National Statistics</i> .	16
2.8.	Evolución de las tasas de mortalidad por suicidio de mujeres, considerando las edades agrupadas en distintas franjas. Fuente: <i>Office for National Statistics</i> .	17
4.1.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.sums</i> .	68
4.2.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.sparsity</i> .	69
4.3.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.within</i> para las respuestas.	71
4.4.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.within</i> para las dosis.	72
4.5.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.within</i> para los ratios de mortalidad.	73
4.6.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.level</i> para las respuestas.	74
4.7.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.level</i> para las dosis.	74
4.8.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.data.level</i> para los ratios de mortalidad.	75
4.9.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.fit.pt</i> .	78
4.10.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.fit</i> escogiendo la opción <i>detrend</i> .	80
4.11.	Resultados gráficos obtenidos al ejecutar la función <i>apc.plot.fit</i> escogiendo la opción <i>sum.sum</i> .	81
5.1.	Representación de la suma de los datos en función de los índices edad, periodo y cohorte.	87
5.2.	Gráfico <i>sparsity</i> para los suicidios (<i>responses</i>) y la población (<i>doses</i>).	88
5.3.	Representación de las cifras de suicidios totales (<i>responses</i>) teniendo en cuenta, únicamente, dos índices temporales.	89
5.4.	Representación de la población (<i>dose</i>) teniendo en cuenta, únicamente, dos índices temporales.	90
5.5.	Representación de las tasas de mortalidad por suicidio (<i>rates</i>) teniendo en cuenta, únicamente, dos índices temporales.	91
5.6.	Representación de las estimaciones de los parámetros haciendo uso de la metodología <i>detrend</i> .	93
5.7.	Mapa de ajuste de la probabilidad.	94
5.8.	Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las cifras de suicidio.	95
5.9.	Representación gráfica de los valores ajustados obtenidos al implementar un modelo edad- periodo- cohorte aplicado a las cifras de suicidio.	95

5.10. Representación gráfica de los predictores lineales obtenidos al implementar un modelo edad- periodo- cohorte a las cifras de suicidio.	96
5.11. Representación gráfica de las respuestas (<i>responses</i>).	96
5.12. Representación gráfica de las tasas de mortalidad por suicidio (<i>rates</i>).	96
5.13. Suma de los datos en función de las variables edad, periodo y cohorte.	99
5.14. Gráfico tipo <i>sparcity</i> para las tasas de mortalidad por suicidio.	100
5.15. Gráfica de las tasas de mortalidad por suicidio utilizando, únicamente, dos índices temporales.	101
5.16. Distribución de las tasas de mortalidad por suicidio en función de las variables edad y periodo.	102
5.17. Mapa de ajuste de probabilidad.	104
5.18. Representación de las estimaciones de los parámetros haciendo uso de la metodología <i>detrended</i> .	105
5.19. Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio.	105
5.20. Transformadas de probabilidad obtenidas al aplicar un modelo edad- cohorte a las tasas de mortalidad por suicidio.	108
5.21. Representación gráfica de los residuos obtenidos al implementar un modelo edad- cohorte a las tasas de mortalidad por suicidio.	109
5.22. Suma de los datos en función de las variables edad, periodo y cohorte; segmentando, a su vez, por sexos.	112
5.23. Gráficos tipo <i>sparcity</i> para las tasas de mortalidad por suicidio de varones y mujeres.	113
5.24. Distribución de las tasas de mortalidad por suicidio en función de las variables edad y periodo.	114
5.25. Gráficos de las tasas de mortalidad por suicidio utilizando dos índices temporales.	115
5.26. Representación de las transformadas de probabilidad diferenciando por sexos.	118
5.27. Representación gráfica de las estimaciones de los parámetros haciendo uso de la metodología <i>detrended</i>	119
5.28. Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio.	121

Índice de cuadros

2.1. Suicide mortality for periods and age groups (5- year intervals), according to sex, England and Wales, 1985 to 2019. Source: <i>Office for National Statistics</i>	26
2.2. Suicide mortality rates, per 100,000 inhabitants, for periods and age groups (5- year intervals), according to sex, England and Wales, 1985 to 2019. Source: <i>Office for National Statistics</i> . . .	27
3.1. Grados de libertad de los distintos submodelos asociados a los modelos edad- periodo- cohorte. Fuente: Nielsen (2015) [129].	53

Capítulo 1

Introducción

El suicidio es considerado como uno de los principales problemas de salud pública existentes en los países desarrollados. De tal forma, en Inglaterra y Gales se produjeron, en 2019, una media de 18 suicidios por día (*Office for National Statistics*), cifra que ilustra la importancia de dicho problema. Además, el suicidio no causa, únicamente, una pérdida irreparable sobre el sujeto que lo comete; ya que, adicionalmente, genera una serie de costes sociales y económicos. Ante esta situación, se ha desarrollado una creciente literatura que ha tratado de estudiar las cuestiones estadísticas que subyacen a los patrones de suicidio, con el objetivo de aportar evidencia científica que sustente el diseño de políticas sanitarias encaminadas a reducir la prevalencia de dicha enfermedad.

De tal forma, en el estudio empírico de los suicidios se han usado, en términos generales, distintos enfoques; caracterizándose, estos últimos, por modelizar las tasas de mortalidad por suicidio, en lugar de las cifras brutas. Además, las diferencias existentes entre los patrones de suicidio de hombres y mujeres han generado que, dichos análisis, hayan sido desarrollados a través de segmentaciones realizadas en función del sexo del fallecido.

En cuanto a los enfoques utilizados para estudiar dichas variables, la mayoría de estudios cuantitativos se caracterizan por modelizar los regresores edad, periodo y cohorte; con el objetivo de analizar los efectos que estos últimos poseen en relación con las tasas de mortalidad por suicidio. De tal forma, autores como Bell y Jones (2013) [1] hacen una alusión teórica a los efectos asociados a dichas variables temporales, en relación con las tasas de mortalidad por suicidio, con el objetivo de clarificar, en este contexto, el significado de dichos regresores. En primer lugar, según estos autores, los efectos asociados a la variable edad, podrían estar relacionados con un aumento de las ratios de mortalidad por suicidio a medida que aumenta la edad; mientras que, en relación con la variable cohorte, eventos como conflictos bélicos, que afectan a cohortes de nacimiento determinadas, podrían derivar en aumentos en las tasas de mortalidad por suicidio de aquellas personas que han vivido durante el desarrollo de dichos conflictos. Por otro lado, un efecto asociado a la variable periodo podría estar relacionado con determinados eventos, como crisis económicas o climáticas, que aumenten las cifras de suicidios, con independencia de los valores asociados a las variables edad y cohorte.

De tal forma, partiendo de estas ideas, autores como Skegg y Cox (1991) [2], Kino et al. (2019) [3] o Martínez- Alés et al. (2021) [4] han implementado modelos de tipología edad- periodo- cohorte con el objetivo de estudiar los efectos que poseían, dichos regresores, en relación con las tasas de mortalidad por suicidio de Nueva Zelanda, Corea del Sur- Japón y Estados Unidos, respectivamente. Sin embargo, la relación de multicolinealidad perfecta existente entre los regresores edad, periodo y cohorte, genera un problema de identificación en este tipo de modelos, impidiendo la estimación directa de los efectos lineales asociados a dichos regresores. Para solucionar esta cuestión, se han implementado distintos procedimientos, entre los que destaca, por ejemplo, el método de parametrización canónica propuesto por Kuang et al. (2008) [5]. En este aspecto, dicha metodología se caracteriza por dividir los efectos temporales del modelo en una parte lineal (sujeta al problema de identificación y, por tanto, no estimable de forma única) y en una parte de tipología no lineal (estimable de forma única).

Por otro lado, autores como Keyes y Li (2012) [6] o Légaré y Hamel (2013) [7] han implementado enfoques de tipología multifase; caracterizándose, estos últimos, por la utilización de métodos gráficos y procedimientos analíticos destinados a analizar las tasas de mortalidad por suicidio. En relación con los instrumentos analíticos, estos autores han utilizado la metodología *median polish*, procedimiento que permite aislar los efectos asociados a una de las tres variables independientes del modelo (edad, periodo o cohorte).

Por otra parte, en este estudio se implementará un modelo edad- periodo- cohorte con el objetivo de estudiar las cifras y las ratios de mortalidad por suicidio (por 100000 habitantes) de Inglaterra y Gales, pertenecientes al periodo 1985- 2019. De tal forma, la base de datos que se utilizará en este estudio ha sido extraída de *Office for National Statistics*; poseyendo, esta última, una estructura de *age- period array*. Además, los datos empleados se caracterizan por poseer intervalos de 5 años.

Por otra parte, como instrumento de ejecución de dichos modelos se empleará el software R. En concreto, se utilizará el paquete *apc* (Fannon y Nielsen, 2020) [8], que utiliza la metodología de parametrización canónica propuesta por Kuang et al. (2008) [5]. De tal forma, este paquete permitirá modelizar las cifras brutas de suicidio empleando la familia de modelos Poisson; considerando la cifra de población total, a su vez, como exposición. Por otro lado, para modelizar las ratios de mortalidad por suicidio, se considerará la familia de modelos gaussianas. Además, debido a las importantes divergencias que existen, en materia de suicidios, entre hombres y mujeres, los resultados serán segmentados por sexo, con el objetivo de realizar actividades de comparación. Por otro lado, para completar el análisis, se utilizarán métodos gráficos que permitan complementar los resultados que han sido obtenidos de forma analítica. Dichos gráficos serán implementados, igualmente, en base al paquete *apc*.

Tras esta breve introducción, el documento se estructurará de la siguiente forma. En primer lugar, en el próximo apartado, se realizará una revisión de las investigaciones que se han desarrollado, en materia de suicidios, considerando la metodología de los modelos edad- periodo- cohorte. Además, se describirán las principales características referentes a los suicidios que se producen en Inglaterra y Gales; considerando, además, las cuestiones inherentes a la base de datos que se utilizará en el proyecto. Posteriormente, en el siguiente epígrafe, se estudiarán los fundamentos teóricos inherentes a los modelos edad- periodo- cohorte; incidiendo, a su vez, en el problema de identificación asociado a dichos modelos. En la siguiente sección, se introducirán las funcionalidades pertenecientes al paquete *apc*; utilizando, para ello, una de las bases de datos que vienen implementadas, de forma predeterminada, en dicho paquete. Posteriormente, en el siguiente epígrafe del estudio, se aplicará un modelo edad- periodo- cohorte a las cifras brutas y las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019). Además, se discutirán los resultados obtenidos en este último apartado; comparando estos, a su vez, con las conclusiones planteadas por otros autores. Finalmente, en el último apartado del proyecto, se establecerán las principales conclusiones obtenidas; realizando una revisión, además, de las principales limitaciones asociadas al estudio.

Capítulo 2

Motivación

El suicidio es considerado como uno de los principales problemas de salud pública existentes en Europa, poseyendo una tasa de prevalencia de 11,93 por 100000 habitantes, según cifras de la *Confederación de Salud Mental española* (en España, esta tasa es del 7,79%). Desde un punto de vista global, según cifras de la OMS (*Organización Mundial de la Salud*), unas 800000 personas se suicidan anualmente; aunque, este problema es especialmente preocupante en los países desarrollados¹. De hecho, en Estados Unidos, país en el que se suicidan anualmente unos 50000 individuos, el suicidio es la cuarta causa de mortalidad. En este aspecto, si se consideran, únicamente, las muertes relacionadas con circunstancias violentas, el suicidio es la principal causa de mortalidad del país americano. Estas cuestiones también se repiten en Canadá, donde el suicidio es la principal causa de mortalidad relacionada con muertes asociadas con traumatismos. De hecho, este problema es especialmente preocupante en el estado de Quebec, donde las muertes por suicidio doblan a los fallecimientos causados por accidente de tráfico. En Europa, existen importantes diferencias geográficas en relación con las tasas de mortalidad por suicidio; ya que, los países escandinavos y del centro de Europa poseen ratios de mortalidad por suicidio superiores a las que presentan las regiones del mediterráneo (Diekstra, 1993 [10]; La Vecchia et al., 1994 [11]). Por otro lado, en países desarrollados asiáticos como Japón o Corea del Sur, se observan, incluso, tasas de mortalidad por suicidio superiores a las de la media de la OCDE (Kim et al., 2011) [12]. De hecho, Corea del Sur presentó la ratio de mortalidad por suicidio más alta de la OCDE en 2019 (24.6 por 100000 habitantes)². Por otro lado, la magnitud del problema de los suicidios se encuentra subestimado, debido a cuestiones legales, sociales, médicas y económicas, como señalan autores como Sainsbury y Jenkins (1982) [13] o Moens (1985) [14]. Por otra parte, en relación con los factores de riesgo que inciden en los suicidios destacan: la edad, el género (el segmento de los varones es más propenso a cometer actos de suicidio) o la salud mental del sujeto, según las investigaciones realizadas por autores como Woodbury et al. (1988) [15] o Bille-Brahe (1993) [16]. En este aspecto, la pirámide poblacional de las sociedades desarrolladas, caracterizada por un progresivo aumento de la esperanza de vida, unido a la prevalencia de ciertas enfermedades crónicas, ha incrementado el número de suicidios en las capas sociales con mayor edad (Woodbury et al., 1988) [15]. Por otro lado, en cuanto a adolescentes y adultos jóvenes, el suicidio es una de las principales causas de fallecimiento; poseyendo asociada, además, una tendencia creciente. En cuanto a los factores de riesgo asociados a los suicidios en estos grupos de edad, destacan: la proliferación de enfermedades mentales (como la depresión), el consumo excesivo de alcohol y drogas o los intentos de suicidio realizados con anterioridad (Moens et al., 1988) [17]. Otros factores como el ambiente familiar, el desempleo o el creciente contagio de enfermedades de transmisión sexual, también pueden haber motivado este incremento de las tasas de mortalidad por suicidio entre jóvenes (Coté et al., 1992) [18]. Además, en términos generales, las diferencias e inconsistencias asociadas a los sistemas de registro y reporte de suicidios, dificultan el cálculo y la interpretación de tendencias asociadas a estas prácticas, sobre todo en relación con el análisis de la prevalencia y de los factores de riesgo asociados a los suicidios. Por otro lado, las cifras oficiales de suicidios se encuentran subestimadas; debido, principalmente, a los diferentes sistemas de determinación de las causas de mortalidad existentes en los distintos países, como señalan autores como Värnik et al. (2010) [19] o Rockett et al. (2011) [20]. De hecho, Tøllefsen et al. (2012) [21] obtuvieron que,

¹En este aspecto, a nivel global, las tasas de mortalidad por suicidio poseen una tendencia decreciente desde hace tres décadas, como señalan autores como Naghavi (2016) [9].

²Para obtener más información relacionada con las tasas de mortalidad por suicidio de los países de la OCDE, consultar el siguiente enlace: <https://data.oecd.org/healthstat/suicide-rates.htm>

en torno a un 10 % de los suicidios cometidos en los países desarrollados, son atribuidos a otras causas de mortalidad. Estas conclusiones también fueron obtenidas por Aoba et al. (2011) [22]; aunque, dichos autores encontraron que, la clasificación errónea del número de suicidios que se produjeron en las metrópolis francesas, durante 2006, era del 9 %; siendo este problema especialmente preocupante en los segmentos de mujeres y personas de avanzada edad. Sin embargo, autores como Charron (1982) [23] y St- Laurent y Bouchard (2004; pp.23) [24] sostienen que, esta subestimación del número de suicidios, no impide que se obtengan resultados válidos, desde el punto de vista de la obtención de conclusiones que sean representativas.

Por otra parte, el suicidio no solo causa una pérdida irreparable en el fallecido, sino que también genera una serie de costes sociales y económicos. Por tanto, la globalidad de este fenómeno, unido a los costes económicos y humanos asociados al mismo, han generado una creciente literatura enfocada a analizar los suicidios. En este aspecto, el objetivo de este estudio consiste en analizar las cifras brutas y las tasas de mortalidad por suicidio de Inglaterra y Gales, pertenecientes al periodo 1985- 2019. Para tal fin, se empleará un modelo edad- periodo- cohorte.

Tras realizar esta breve introducción del tema de estudio, se procederá a describir, de forma detallada, las problemáticas más destacadas que existen, en Inglaterra y Gales, en relación con el suicidio. Posteriormente, se introducirá una breve revisión de la literatura relacionada con el estudio de los suicidios. En este aspecto, se hará referencia, únicamente, a los estudios que han utilizado una perspectiva estadística; ya que, los proyectos relacionados con cuestiones cualitativas, quedan fuera del alcance de este estudio. Finalmente, se describirá la base de datos utilizada en el análisis.

2.1. El problema de los suicidios en Inglaterra y Gales

Desde un punto de vista general, el suicidio es uno de los principales problemas de salud pública existentes en Reino Unido. En este aspecto, en el año 2019, se registraron 5691 suicidios en Inglaterra y Gales (11 muertes por 100000 habitantes); por lo que, hubo una media de 18 suicidios por día. De hecho, en Inglaterra y Gales, el suicidio fue la principal causa de mortalidad en varones menores de 45 años. En este aspecto, las tendencias suicidas son especialmente preocupantes en varones; ya que, en 2019, en torno al 75 % de los suicidios fueron cometidos por hombres. En este aspecto, en la figura 2.1 se introduce la evolución de las tasas de mortalidad por suicidio, por 100000 habitantes, correspondientes a la población mayor de 10 años que habita en Inglaterra y Gales. En este aspecto, existe una importante disminución de las tasas de mortalidad por suicidios, tanto en varones como en mujeres, entre 1981 y 2019; aunque, en los últimos años de la serie, se observa un ligero incremento de dicha ratio.

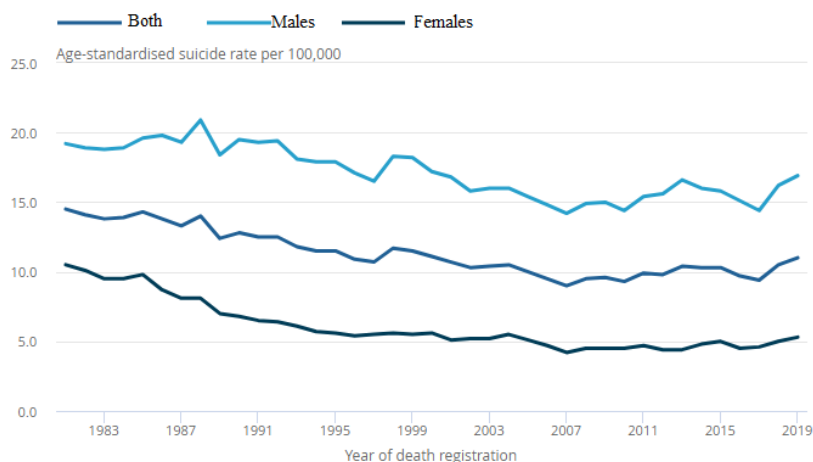


Figura 2.1: Ratios de mortalidad por suicidio, registradas en Inglaterra y Gales, entre 1981 y 2019. Se considera, únicamente, a la población mayor de 10 años. Fuente: *Office for National Statistics*.

Si estas cifras son desagregadas por países, Inglaterra poseía, en 2019, una ratio de mortalidad por suicidio de 10,8 por 100000 habitantes. En este aspecto, fallecieron por suicidio, concretamente, 5316 personas; siendo, el 75,56 % de estos individuos varones. La evolución de las tasas de mortalidad por suicidio (por 100000 habitantes) se presenta, para el caso de Inglaterra, en la figura 2.2. De tal forma, se observa una disminución destacada de las tasas de mortalidad por suicidio entre 1981 y 2019, tanto en el caso de varones como en el segmento de mujeres.

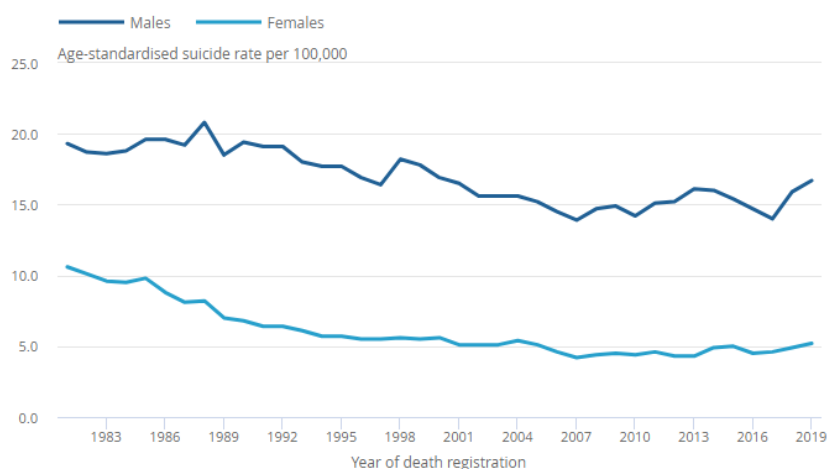


Figura 2.2: Ratios de mortalidad por suicidio registradas en Inglaterra entre 1981 y 2019; considerando, únicamente, a la población mayor de 10 años. Fuente: *Office for National Statistics*.

En el caso de Gales, se observa una importante volatilidad en las tasas de mortalidad por suicidio de varones, como se observa en la figura 2.3. Esta situación está provocada por el reducido número de suicidios que ocurren en dicho país (en concreto, en 2019, se suicidaron 248 varones y 82 mujeres).

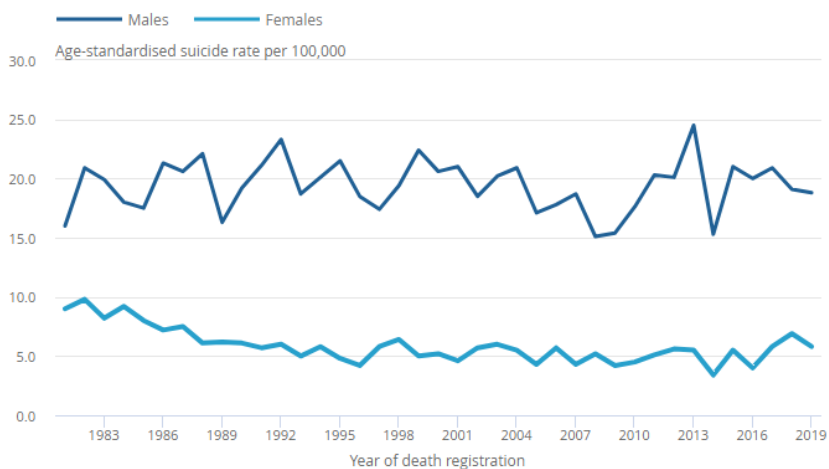


Figura 2.3: Ratios de muertes por suicidio, registradas en Gales, entre 1981 y 2019. Fuente: *Office for National Statistics*.

Por otro lado, existen importantes diferencias regionales en relación con las tasas de mortalidad por suicidio, cuestión que debería ser analizada en posteriores investigaciones. En este aspecto, considerando las tasas de mortalidad por suicidio de varones, la región de Yorkshire y The Humber presenta las mayores tasas de mortalidad por suicidio de Inglaterra y Gales (20,6 muertes por 100000 habitantes en 2019). Por otro

lado, en la región de South East, las tasas de mortalidad por suicidio han crecido, de forma importante, durante los últimos años. Para ilustrar las diferencias regionales referentes a las ratios de mortalidad por suicidio; se introducen, en la figura 2.4, las tasas de mortalidad por suicidio de varones (por 100000 habitantes) pertenecientes a las distintas regiones de Inglaterra y Gales.

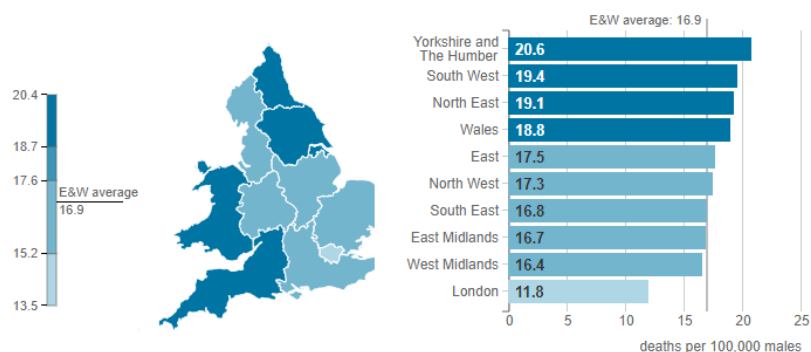


Figura 2.4: Ratios de mortalidad por suicidios registradas en 2019; considerando, únicamente, los fallecimientos de varones. Fuente: *Office for National Statistics*.

En el caso del colectivo femenino, la región de Yorkshire y The Humber poseía, en 2019, la tasa de mortalidad por suicidio más elevada (7,3 muertes por suicidio por 100000 habitantes). Para ilustrar las diferencias regionales existentes en relación con las tasas de mortalidad por suicidio; se introducen, en la figura 2.5, las ratios de mortalidad por suicidio de mujeres (por 100000 habitantes), considerando las distintas regiones que componen Inglaterra y Gales. Por otro lado, existen importantes diferencias entre las ratios de mortalidad por suicidio de varones y mujeres; por lo que, los datos serán segmentados, en el análisis, en función del sexo del individuo.

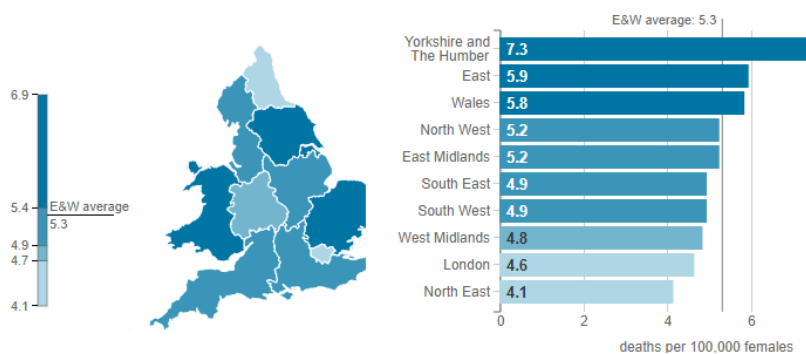


Figura 2.5: Ratios de mortalidad por suicidio registradas en 2019; considerando, únicamente, los fallecimientos de mujeres. Fuente: *Office for National Statistics*.

Para hacer frente a este importante problema de salud pública, las instituciones británicas han desarrollado diferentes iniciativas enfocadas a prevenir el suicidio, entre las que destacan: la creación de una subsecretaría gubernamental (*Parliamentary Under Secretary of State for Mental Health, Inequalities and Suicide Prevention*) y la puesta en marcha de un teléfono de prevención del suicidio.

2.1.1. Definición de suicidio

En base a la definición de suicidio planteada por *Office for National Statistics*, las cifras y las tasas de mortalidad por suicidio con las que se trabajará en este estudio harán referencia a una de las siguientes

cuestiones:

- Fallecimiento causado por una autolesión realizada de forma intencionada. Debe ser dictaminado por un médico forense, una vez realizada la pertinente autopsia.
- Fallecimiento causado por un evento de intención indeterminado; que ha sido considerado, por un médico forense, como un suicidio (es decir, en base a este profesional, la probabilidad de que la muerte haya estado asociada a un suicidio es alta).

Por tanto, según la institución estadística británica, esta definición de suicidio es distinta a la utilizada por otros países; por lo que, en Inglaterra y Gales, se produce una sobreestimación de la mortalidad por suicidio, en comparación con algunos países.

Por otra parte, todos los posibles fallecimientos asociados a los suicidios deben ser confirmados, en Inglaterra y Gales, por un médico forense; siguiendo, a su vez, un estricto procedimiento. De tal forma, la notificación del hipotético suicidio no se realizaría hasta finalizar dicha investigación; pudiendo extenderse, esta última etapa, durante meses (o incluso años). De hecho, según *Office for National Statistics*, únicamente el 51 % de los suicidios que se registraron en 2019 ocurrieron en el citado año. Si se analiza el valor mediano referente a los retrasos en la notificación de los suicidios, en lugar de la media (ya que, esta medida está afectada por casos extremos), dicho estadístico tomó un valor, en 2019, de 166 días (considerando el caso de Inglaterra); mientras que, en Gales, tomó un valor de 165 días.

De tal forma, las estadísticas utilizadas en este proyecto están basadas en el año de registro del suicidio (no en el año en el que este se produjo, como bien se acaba de explicar). Por tanto, esta situación dificulta las tareas de análisis de tendencia, sobre todo en aquellas regiones que poseen una población escasa. Este es, por tanto, uno de los principales problemas asociados a esta base de datos.

2.1.2. Patrones de suicidio en función de la variable edad

Es interesante analizar los patrones de suicidio en función de la edad de la víctima; ya que, esto posibilitará la detección de las características asociadas a los suicidios. De tal forma, a partir de la década de los ochenta del siglo XX, se observa un desplazamiento de la edad en la que se produce el pico de suicidios referente a las personas de mediana edad, pasando de los 45 años a los 54. Tras alcanzar este máximo, se observa un decrecimiento de las tasas de mortalidad por suicidio. Finalmente, dichas ratios vuelven a incrementarse a partir de los 80 años (*Office for National Statistics*).

Por otro lado, existen importantes diferencias respecto a la edad en la que, las tasas de mortalidad por suicidio de varones y mujeres, alcanzan el máximo global. De tal forma, en el segmento de varones, dicho máximo se alcanzó, en 2019, entre los 45 y los 49 años; mientras que, en el segmento de mujeres, el máximo global se situó en el grupo de edad comprendido entre los 50 y los 54 años.

Por otra parte, las tasas de mortalidad por suicidio tienden a incrementarse en edades avanzadas, independientemente del sexo, como se observa en la figura 2.6. Esta circunstancia es característica de los países desarrollados y está provocada, a su vez, por un importante número de factores, entre los que destacan los siguientes:

- La proliferación de enfermedades psiquiátricas en estas edades.
- El deterioro físico asociado a personas de avanzada edad.
- Otros factores de tipo social; entre los que destaca, principalmente, la soledad sufrida por un importante número de personas de avanzada edad.

Por otro lado, las tasas de mortalidad por suicidio de los varones con edades comprendidas entre los 10 y los 24 años, se han caracterizado por poseer una escasa cuantía, en comparación con el resto de grupos de edad. Sin embargo, estas tasas han poseído, durante el último lustro, una tendencia creciente. Por otra parte, entre 1995 y 2009, las tasas de mortalidad por suicidio más elevadas se situaban, en el segmento de

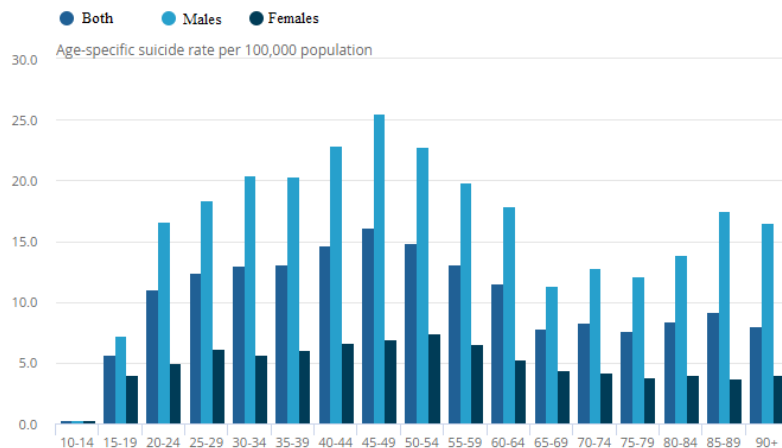


Figura 2.6: Ratios de mortalidad por suicidio considerando la variable edad agrupada en intervalos de 5 años. Datos recolectados en Inglaterra y Galés para el año 2019. Fuente: *Office for National Statistics*.

los varones, entre los 25 y los 44 años. Sin embargo, a partir de 2013, las ratios de mortalidad por suicidio más elevadas, se han concentrado en el grupo de personas que poseen entre 45 y 64 años. En cuanto a las razones que subyacen a las elevadas tasas de mortalidad por suicidio que poseen los varones de mediana edad, destacan: la mayor propensión de estas personas a sufrir los efectos adversos de las crisis económicas y los problemas de alcoholismo y soledad existentes en este grupo de edad. Además, los varones de mediana edad presentan una menor propensión a pedir ayuda de naturaleza psicológica, situación que puede influir, negativamente, en el suicidio. Por otro lado, entre 1981 y 1990, las tasas de mortalidad por suicidio más altas, en el segmento de los varones, pertenecían a personas con 75 o más años. Sin embargo, a partir de 1990, las tasas de mortalidad por suicidio asociadas a este grupo de edad, se han reducido progresivamente. Dichas características pueden observarse en la figura 2.7.

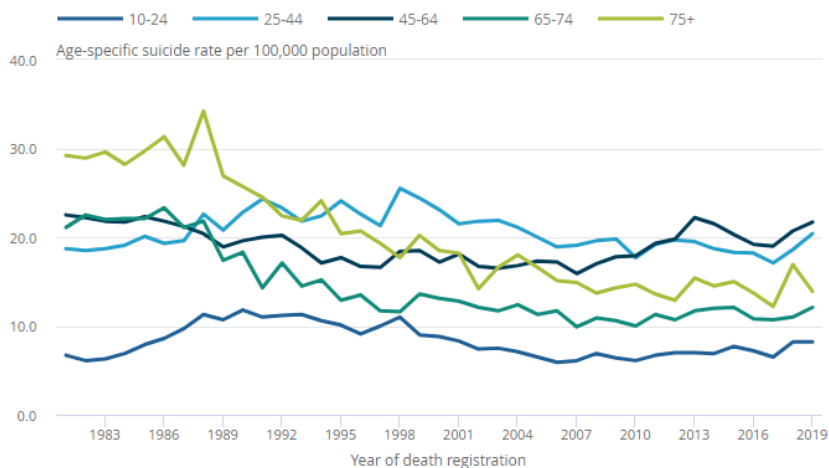


Figura 2.7: Evolución de las tasas de mortalidad por suicidio en el segmento de varones, considerando las edades agrupadas en distintas franjas. Fuente: *Office for National Statistics*.

Por otra parte, en el caso del colectivo femenino, las tasas de mortalidad por suicidio son reducidas, en comparación con las ratios de mortalidad por suicidio que presentan los varones, como se observa en la figura 2.8. En este aspecto, desde 2012, las ratios de mortalidad por suicidio de las mujeres con edades comprendidas entre los 10 y los 24 años, se han visto incrementadas. Por otro lado, entre 2008 y 2016, las tasas de mortalidad por suicidio de mujeres con edades comprendidas entre los 25 y los 44 años, se han visto

estabilizadas; aunque, entre 2007 y 2017, estas se incrementaron ligeramente. Finalmente, entre 1981 y 2019, se han reducido las tasas de mortalidad por suicidio, considerando el segmento de mujeres que poseen edades superiores a los 45 años.

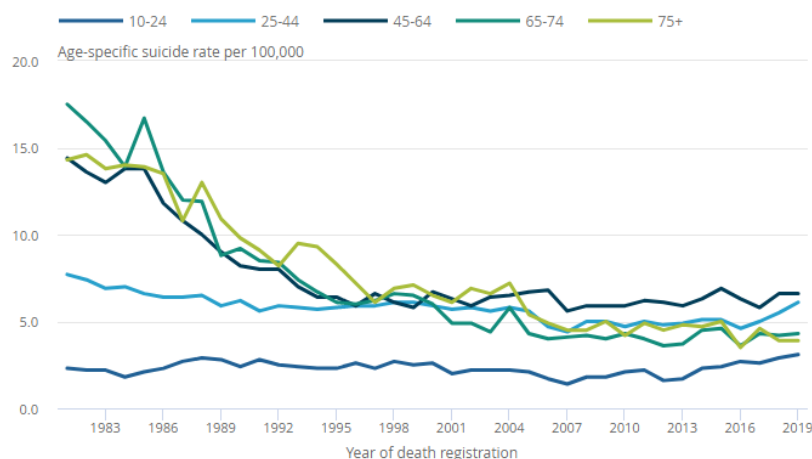


Figura 2.8: Evolución de las tasas de mortalidad por suicidio de mujeres, considerando las edades agrupadas en distintas franjas. Fuente: *Office for National Statistics*.

2.1.3. Otras características de los suicidios

Como se explicó anteriormente, existen una serie de cuestiones que justifican las elevadas tasas de mortalidad por suicidio existentes en Inglaterra y Gales; destacando, en este aspecto: las recesiones económicas, el desempleo, la pobreza o la soledad. De hecho, según *Samaritans*³ las características socio- económicas de los individuos determinan su propensión al suicidio.

Por otro lado, el 61,7% de los varones y el 46,7% de las mujeres utilizaron el ahorcamiento, estrangulamiento y posterior asfixia, en 2019, como procedimiento de suicidio; siendo, este último, el principal método de suicidio empleado desde el comienzo de la serie temporal. La segunda metodología de suicidio más empleada es el envenenamiento; mientras que, los suicidios ejecutados mediante la utilización de armas de fuego son minoritarios, debido a la legislación restrictiva que existe, en Reino Unido, en relación con la tenencia de dichos objetos. Por otra parte, la inhalación de gas doméstico (este contenía partículas de monóxido de carbono, al ser un derivado del carbón) fue uno de los principales métodos de suicidio hasta su sustitución, en la década de los noventa del siglo XX, por gas natural (no venenoso). Posteriormente, la introducción de los catalizadores en los motores de los automóviles permitió que, los suicidios instrumentados a través de la inhalación del monóxido de carbono producido a través de la combustión interna de los vehículos, desapareciesen. En este aspecto, según las investigaciones realizadas por autores como Snowdon y Hunt (2002) [25], Biddle et al. (2008) [26], Hawton y Van Heeringen (2009) [27], Gagné et al. (2010) [28] o Hanigan et al. (2012) [29], la reducción del acceso a posibles métodos de suicidio, como las armas de fuego o el monóxido de carbono, han generado una reducción en las tasas de mortalidad por suicidio.

2.2. Revisión de la literatura

La literatura que aborda el estudio de los patrones de suicidio desde una vertiente cuantitativa está basada, mayoritariamente, en la aplicación de modelos edad- periodo- cohorte (*APC models*)⁴. De tal forma,

³Organización británica sin ánimo de lucro que ofrece apoyo emocional a cualquier persona que posea: estrés, problemas transitorios de carácter mental, riesgo de suicidio... Su área de influencia principal se centra en Reino Unido e Irlanda; donde posee, incluso, líneas de atención telefónicas.

⁴La base teórica de estos modelos será presentada en el capítulo 3.

esta metodología permite estudiar los efectos que las variables edad, periodo y cohorte poseen en relación con los patrones de suicidio.

En primer lugar, se procederá a explicar la definición de estas tres variables en el contexto del estudio de los suicidios, justificando su análisis individualizado a través de la implementación de un modelo edad-periodo-cohorte. Para ello, se utilizará la explicación propuesta por Bell y Jones (2013) [1], que justifica la aplicación de este tipo de modelos a la hora de estudiar los efectos temporales asociados a las tasas de mortalidad por suicidio. En este aspecto, un efecto asociado a la variable edad, podría estar relacionado con un aumento de la propensión al suicidio a medida que la edad del sujeto se ve incrementada. Por otro lado, los efectos asociados a la variable cohorte podrían estar relacionados con una serie de eventos que han sido vividos, de manera simultánea, por un conjunto de personas. Por poner un ejemplo, una cohorte que haya experimentado una guerra, podría poseer una mayor propensión al suicidio (debido a los efectos psicológicos derivados de dichas situaciones), en comparación con cohortes que no hayan estado sometidas a tales periodos de violencia. Finalmente, un efecto asociado a la variable periodo podría estar relacionado con eventos determinados que deriven en una variación de la propensión al suicidio, independientemente de la edad del sujeto o de la cohorte de nacimiento de este. En este aspecto, por ejemplo, una crisis económica podría derivar en un incremento de las tasas de mortalidad por suicidio, independientemente de las variables edad y cohorte⁵.

Una vez realizada esta breve explicación de las variables que conforman los modelos edad-periodo-cohorte en el contexto de los suicidios, se realizará una exposición de los estudios más destacados que se han realizado en esta materia, con el objetivo de extraer conclusiones que puedan servir como elemento de comparación a la hora de analizar las tasas de mortalidad por suicidio (y las cifras brutas de los mismos) que se han producido en Inglaterra y Gales entre 1981 y 2015.

En primer lugar, Granizo et al. (1996) [31] utilizaron un modelo Poisson con el objetivo de estudiar los efectos de las variables edad, periodo y cohorte respecto a las tasas de mortalidad por suicidio españolas (periodo 1959- 1991). Desde un punto de vista teórico, estos autores se basaron en los trabajos de Kupper et al. (1985) [32] y Fromme y Checkoway (1985) [33], autores que analizaron las ventajas de los modelos Poisson a la hora de estudiar tasas. Respecto a las principales conclusiones obtenidas por Granizo et al. (1996) [31], se observaba un aumento significativo de las tasas de mortalidad por suicidio españolas en relación con la variable edad. Además, estos autores obtuvieron que, el efecto del regresor edad, no presentaba diferencias significativas entre hombres y mujeres. Por otro lado, el efecto de la variable periodo resultaba ser significativo; aunque, presentaba una importancia menor, en comparación con el regresor edad, a la hora de explicar las tasas de mortalidad por suicidio. Por otra parte, respecto a los efectos de la variable cohorte, estos eran especialmente importantes en las cohortes nacidas después de 1950, con independencia del sexo. En este aspecto, si se compara la cohorte de referencia (1885- 1889) con las cohortes nacidas a finales de los años sesenta y principios de los setenta del siglo XX; se observa que, estas últimas cohortes, poseen el triple de riesgo de fallecer por suicidio.

De tal forma, en base al estudio realizado por Granizo et al. (1996) [31], los principales determinantes de las tasas de mortalidad por suicidio españolas (periodo 1959- 1991) eran: el género, la edad del sujeto y el periodo. Por otra parte, en relación con los efectos del regresor cohorte, estos resultaban ser estadísticamente significativos en las cohortes de nacimiento más recientes, con independencia del sexo.

Por otro lado, autores como Murphy y Wetzel (1980) [34], Häfner y Schmidtke (1985) [35] y Reed et al. (1985) [36]; que estudiaron, respectivamente, las tasas de mortalidad por suicidio de Estados Unidos (periodo 1949- 1974), Alemania del Oeste (periodo 1951- 1980) y Canadá (periodo 1921- 1980), encontraron un importante efecto de la variable periodo en el segmento de varones. Además, dichos autores hallaron un efecto del regresor cohorte en las cohortes de nacimiento de varones jóvenes. Por otro lado, La Vecchia et al. (1986) [37] y Murphy et al. (1986) [38] encontraron, únicamente, un importante efecto de la variable cohorte; en relación, este último, con las cohortes de nacimiento de varones más antiguas. Dichos estudios se enfocaron, respectivamente, en las tasas de mortalidad por suicidio de Italia (periodo 1955- 1979) e Inglaterra y

⁵Por poner un ejemplo, durante la crisis de las economías asiáticas de 1997, se produjo un notable incremento de las tasas de mortalidad por suicidio en países como Japón, Corea del Sur o Taiwán (Chang et al., 2009) [30].

Gales (periodo 1921- 1980). Por otra parte, Moens et al. (1987) [39] y Åsgård et al. (1987) [40] encontraron, en Bélgica (periodo 1954- 1981) y Suecia (periodo 1952- 1981), respectivamente, un importante efecto del regresor cohorte en las cohortes de nacimiento más jóvenes, con independencia del sexo.

Por otro lado, conviene tener en cuenta que, realizar comparaciones entre los efectos de las variables edad, periodo y cohorte, para distintos años o regiones, reviste de una notable dificultad, debido a:

- La dependencia lineal existente entre los regresores temporales.
- Los cambios que se producen en los sistemas de conteo nacionales; además de, las diferencias existentes, entre los distintos países, en relación con la definición del suicidio.
- Las dificultades que existen a la hora de determinar los factores de riesgo asociados a los suicidios.

Por ende, se debe ser muy cauto a la hora de realizar tales comparaciones, como señalan autores como Granizo et al. (1996) [31].

Por otro lado, Légaré y Hamel (2013) [7] realizaron un análisis edad- periodo- cohorte con el objetivo de estudiar las tasas de mortalidad por suicidio de la provincia de Quebec (Canadá) durante el periodo 1950- 2009. Esta región del norte de América es especialmente interesante desde el punto de vista de las pautas de suicidio de varones; ya que, fue, durante décadas, la región de Canadá que presentaba las mayores ratios de mortalidad por suicidio (recientemente, ha sido superada por las provincias de Yukon y Nuvavut); reduciéndose dichas tasas de mortalidad, además, con la edad (a diferencia de la mayoría de países desarrollados, donde las tasas de mortalidad por suicidio se incrementan con la edad). En este aspecto, autores como Beaupré y St- Laurent (1998) [41] sospechan que existe un efecto de cohorte, en relación con las tasas de mortalidad por suicidio, entre los individuos nacidos tras la 2ª Guerra Mundial (generación del *baby boom* canadiense)⁶.

Por otra parte, las diferencias existentes entre las tasas de mortalidad por suicidio de hombres y mujeres motivaron que, el estudio realizado por Légaré y Hamel (2013) [7], se desarrollase siguiendo una segmentación en función del sexo, práctica habitual que ha sido implementada por autores como Barnes et al. (1986) [43], Mao et al. (1990) [44] Beaupré y St- Laurent (1998) [41] o Keyes y Li (2012) [6]; y que, por tanto, será aplicada en este proyecto. Partiendo de dicha segmentación, Légaré y Hamel (2013) [7] emplearon un análisis edad- periodo- cohorte basado en una aproximación multifase, en la que combinaron tanto métodos gráficos como procedimientos estadísticos. El objetivo de dicha metodología consistía en aislar el efecto de la variable cohorte; ya que, dichos autores utilizaron un *age- period array*. De tal forma, la metodología gráfica pretendía analizar aquellas variaciones en las tasas de mortalidad por suicidio que se encontraban asociadas a las variables periodo y cohorte; mientras que, el método estadístico, se basaba en aplicar un *median polish* a la transformación logarítmica de las ratios de mortalidad por suicidio. De tal forma, el *median polish* permitía eliminar los efectos aditivos asociados a las variables edad y periodo, permitiendo aislar los efectos del regresor cohorte (Keyes et al., 2010) [45]. Para lograr el aislamiento de dicha variable, Légaré y Hamel (2013) [7] siguieron los siguientes pasos:

1. Considerando la tabla en la que estaban contenidas las tasas de mortalidad por suicidio; estos autores extrajeron, en primer lugar, las medianas correspondientes a cada fila y columna.
2. Tras realizar algunas iteraciones, se obtuvo una matriz de residuos; formada, a su vez, por dos componentes: los efectos aislados de la variable cohorte (*systematic components*) y una serie de errores aleatorios (*non- systematic components*).
3. Aplicación de un modelo de regresión lineal a dichos residuos aleatorios. De esta manera, se obtuvieron los efectos estimados de cada cohorte; que presentaban, a su vez, forma de ratios (con sus correspondientes categorías de referencia).

⁶Esta circunstancia no es exclusiva de Canadá; ya que, autores como Phillips (2014) [42] han hallado evidencias que sostienen que, la generación del *baby boom* estadounidense (personas nacidas tras concluir la 2ª Guerra Mundial; es decir, nacidas entre 1946- 1964), poseen tasas de suicidios superiores a las de generaciones posteriores.

Desde el punto de vista del método gráfico, las conclusiones obtenidas por estos autores son especialmente interesantes, pudiendo utilizarse como elemento comparativo en este proyecto. En primer lugar, las tasas de mortalidad por suicidio de varones mostraban un incremento con la edad (partiendo de la adolescencia); hasta alcanzar, finalizando la cuarentena, el máximo global de la serie. Sin embargo, dichas tasas se han visto sometidas a cambios notables a lo largo del tiempo, como también señalan autores como Boyer y St-Laurent (1999) [46]. Por otro lado, las ratios de mortalidad por suicidio de mujeres presentaban un comportamiento estable, en relación con la variable edad, a lo largo del periodo de análisis. Por otra parte, las cohortes nacidas entre 1945 y finales de la década de los ochenta del siglo XX, presentaban tasas de mortalidad por suicidio superiores a las de la cohorte de referencia.

Por otro lado, la aplicación de la metodología *median polish* permitió obtener resultados interesantes. En primer lugar, las cohortes de varones nacidas entre 1870 y 1909 (con la excepción de las cohortes nacidas entre 1890- 1894) poseían un riesgo de suicidio ligeramente superior respecto a las cohortes de varones nacidas durante el periodo 1940- 1944 (cohorte de referencia). Esta situación también se repetía con las cohortes de varones nacidas entre 1950 y 1979; aunque, el riesgo de suicidio de dichas cohortes era inferior al de las cohortes nacidas entre 1870 y 1909. Por otro lado, en relación con el segmento de mujeres, se observaba un ligero efecto del regresor cohorte; que destacaba, principalmente, en las cohortes nacidas entre los años 1875-1894 y 1980- 1989. En este aspecto, autores como Reed et al. (1985) [36] (estudiaron las ratios de mortalidad por suicidio de Canadá), Barnes et al. (1986) [43] (analizaron las tasas de mortalidad por suicidio de la provincia canadiense de Ontario), Mao et al. (1990) [44] (estudiaron las ratios de mortalidad por suicidio de Canadá) o Beupré y St- Laurent (1998) [41] (analizaron las tasas de mortalidad por suicidio de Quebec); obtuvieron que, los efectos asociados a la variable cohorte, eran inferiores a los de los regresores edad y periodo; siendo, además, más intensos en el segmento de varones. Por tanto, estos autores confirman los resultados obtenidos por Légaré y Hamel (2013) [7].

De tal forma, Légaré y Hamel (2013) [7] encontraron que, los regresores edad y periodo afectaban, en mayor medida que la variable cohorte, a las tasas de mortalidad por suicidio de los habitantes de Quebec. En este aspecto, las variaciones observadas en las ratios de mortalidad por suicidio de mujeres estaban relacionadas, en mayor medida, con la variable edad; mientras que, las fluctuaciones de las tasas de mortalidad por suicidio de varones estaban asociadas, principalmente, con el regresor periodo⁷. Por otro lado, autores como Morrel et al. (2002) [47] y Odagiri et al. (2009) [48]; que analizaron, respectivamente, las tasas de mortalidad por suicidio de Nueva Gales del Sur (Australia) y Japón, también encontraron un menor efecto de la variable cohorte a la hora de explicar los suicidios, en comparación con los efectos de los regresores edad y periodo. Sin embargo, autores como Keyes y Li (2012) [6], que analizaron las ratios de mortalidad por suicidio estadounidenses empleando el método *median polish*; obtuvieron que, los efectos del regresor periodo, resultaban ser no significativos a la hora de explicar dichas tasas. Además, estos autores encontraron evidencias que sustentaban la existencia de un efecto creciente, sobre las ratios de mortalidad por suicidio, asociado a la variable cohorte. En este aspecto, dicho efecto incidía, principalmente, en las cohortes de varones nacidas a partir del comienzo de la década de los ochenta del siglo XX.

Por otra parte, Légaré y Hamel (2013) [7] señalaron que, algunas medidas implementadas por el Gobierno de Canadá (y la provincia de Quebec), han resultado ser exitosas a la hora de reducir las tasas de mortalidad por suicidio; aunque, se necesitaría realizar una investigación en profundidad acerca de los efectos de dichas medidas, según estos autores. De tal forma, entre las estrategias desarrolladas por la provincia de Quebec destaca, por ejemplo, la *Stratégie Québécoise de prévention du suicide*⁸. Por otro lado, en base a los resultados obtenidos por Légaré y Hamel (2013) [7], las estrategias de prevención del suicidio deberían centrarse en factores de riesgo como la edad, en detrimento de las cohortes (debido al escaso peso estadístico de dicha variable, en relación con la explicación de la evolución de las tasas de mortalidad por suicidio).

⁷Estos efectos temporales referentes al colectivo de varones (y a la variable periodo) han sido encontrados, a su vez, en otros países. En este aspecto, por ejemplo, las recesiones económicas y los fenómenos meteorológicos extremos, han sido asociados con aumentos en las tasas de mortalidad por suicidio de varones.

⁸Entre las medidas que incluye esta estrategia destacan, por ejemplo, la detección precoz de posibles suicidas, el tratamiento de depresiones por parte de profesionales, el monitoreo de personas que han intentado suicidarse con anterioridad, la limitación de acceso a posibles fuentes de suicidio para personas con riesgo alto de suicidio o la implantación de programas destinados a prevenir las actividades suicidas entre jóvenes.

Por otro lado, las preocupantes tasas de mortalidad por suicidio existentes en Estados Unidos, sobre todo en las regiones poco pobladas de los estados occidentales no costeros (Stone et al., 2015) [49], han generado un importante número de investigaciones. En este aspecto, recientemente, se han incrementado las ratios de mortalidad por suicidio de las personas con edades comprendidas entre los 10 y los 44 años; además de, entre los varones mayores de 64 años. Por otro lado, la situación de Estados Unidos, en materia de suicidios, presenta notables diferencias en relación con el resto de economías occidentales; debido, principalmente, al laxo control de armas de fuego que existe en dicho país.

En primer lugar, Murphy y Wetzel (1980) [34] hallaron importantes efectos de la variable cohorte, respecto a las tasas de mortalidad por suicidio estadounidenses, en las cohortes nacidas tras la conclusión de la 2ª Guerra Mundial. Por otra parte, Wang et al. (2016) [50] encontraron, empleando las tasas de mortalidad por suicidio estadounidenses pertenecientes al periodo 1983- 2012, diferencias en los efectos asociados a la variable edad entre personas de raza blanca e individuos de raza negra. Por otro lado, Martínez- Alés et al. (2021) [4] utilizaron un modelo edad- periodo- cohorte con el objetivo de estimar los efectos que poseían, dichas variables, en las tasas de mortalidad por suicidio estadounidenses pertenecientes al periodo 1999-2018; segmentando los datos, a su vez, en función del sexo, la raza y el método de suicidio utilizado. En primer lugar, estos autores hallaron que, el modelo general (compuesto por las variables edad, periodo y cohorte), presentaba la mejor capacidad de ajuste, en comparación con el resto de submodelos. Además, hallaron evidencias que sostienen que, las cohortes de los individuos más jóvenes (especialmente las compuestas por adolescentes de grupos étnicos minoritarios), se caracterizaban por ser colectivos con especial riesgo de suicidio. De tal forma, considerando, únicamente, los suicidios perpetrados mediante métodos distintos a las armas de fuego, estos autores encontraron un efecto significativo de la variable cohorte, relacionado con un incremento del riesgo de suicidio, en las comunidades de jóvenes afroamericanos, asiáticos y nativos americanos.

Por otra parte, la posesión de un arma de fuego incrementa el riesgo de suicidio. De hecho, en Estados Unidos, el 60 % de las muertes por armas de fuego se corresponden con suicidios; mientras que, en torno al 50 % de los suicidios, se ejecutan mediante este procedimiento. Sin embargo, durante las últimas décadas, la posesión de armas de fuego se ha visto reducida (Smith y Son, 2019) [51]; incrementándose, por otro lado, el número de suicidios ejecutados mediante otros métodos (asfixia, principalmente). En este aspecto, desde el punto de vista de los suicidios cometidos empleando armas de fuego, se observa un riesgo de suicidio elevado entre jóvenes; además de, entre hombres de raza blanca con edades avanzadas. Sin embargo, Martínez- Alés et al. (2021) [4] encontraron que, las cohortes de jóvenes (tanto de varones como de mujeres), poseían un riesgo de suicidio, relacionado con armas de fuego, cada vez menor; pudiendo estar motivado, este hecho, por los nuevos controles de armas que se han venido aplicando, en Estados Unidos, sobre estos colectivos. Estas conclusiones también han sido planteadas por autores como Ludwig y Cook (2000) [52], Andrés y Hempstead (2011) [53] o Raifman et al. (2020) [54]. Por otro lado, considerando el periodo 1999- 2018 y, desde el punto de vista de la variable periodo, se observa un aumento de los suicidios realizados con métodos distintos a las armas de fuego. De tal forma, considerando este tipo de suicidios, se observa un importante efecto del regresor edad; afectando, principalmente, a los individuos más jóvenes, independientemente de su sexo y raza (excepto para mujeres blancas). En cuanto a los efectos referentes a la variable cohorte (relacionados, estrictamente, con los suicidios ejecutados a través de métodos distintos a las armas de fuego), son especialmente significativos entre jóvenes no blancos; afectando, principalmente, a afroamericanos y asiáticos. Estas conclusiones también fueron obtenidas por autores como Mojtabai et al. (2016) [55], Plemmons et al. (2018) [56] o Keyes et al. (2019) [57]. De tal forma, según Oquendo y Volkow (2018) [58], la razón principal que subyace a este efecto de cohorte es la epidemia de opio que asola Estados Unidos; y que afecta, especialmente, a las cohortes más jóvenes. Por otro lado, si se consideraban, únicamente, los suicidios realizados por personas de raza negra, se observaba un importante efecto de cohorte relacionado con los suicidios ejecutados mediante armas de fuego (Martínez- Alés et al., 2021) [4]. Este efecto resultaba ser especialmente importante en las cohortes de varones nacidas a finales de los años setenta y principios de los años ochenta del siglo XX. En cuanto a los suicidios ejecutados a través de otros métodos distintos a las armas de fuego, se observaba un efecto de cohorte significativo en las cohortes de mujeres negras nacidas al comienzo de la década de los ochenta del siglo XX (Martínez- Alés et al., 2021) [4].

Por otro lado, las elevadas tasas de mortalidad por suicidio existentes en países asiáticos como Corea del Sur o Japón, han sido analizadas en numerosos estudios. En este aspecto, durante los años cincuenta del

siglo XX, Japón alcanzó las tasas de mortalidad por suicidio más elevadas de su historia (Motohashi, 1991) [59]; reduciéndose, dichas cifras, a raíz de la imposición de regulaciones destinadas a controlar las sustancias que eran empleadas, hasta esa fecha, como métodos principales de suicidio (bromuro y gas de carbón). Sin embargo, las cifras de suicidios existentes en Japón siguen siendo elevadas, poseyendo importantes picos situados en los periodos de crisis económica (*asset bubble* en 1991 y *asian economic crisis* en 1997). En este aspecto, las prefecturas con menor nivel económico, presentaban tasas de mortalidad por suicidio superiores a aquellas con un nivel adquisitivo alto, como señalan autores como Suzuki et al. (2014) [60]. Por otra parte, Corea del Sur posee unas ratios de mortalidad por suicidio similares a las presentadas por Japón; debido, principalmente, a las elevadas tasas de mortalidad por suicidio que presentan los ancianos coreanos (Khang et al., 2005) [61]. Además, según autores como Jung- Choi y Kim (2013) [62] o Park et al. (2018) [63], las medidas de austeridad implementadas por el FMI, tras el estallido de la crisis financiera asiática de 1997-1998, unido a un sistema de protección social ineficiente, han generado un aumento en las tasas de mortalidad por suicidio de dicho país. En este aspecto, las ratios de mortalidad por suicidio más elevadas se sitúan en grupos de personas con un nivel socioeconómico reducido (Lee et al., 2009) [64]; además de, en las áreas con menor desarrollo económico (Yoon et al., 2015) [65]. Por otra parte, aunque ambos países han desarrollado políticas encaminadas a reducir las tasas de mortalidad por suicidio, el presupuesto proyectado por Corea del Sur se correspondía, en 2018, con un 2,1 % del presupuesto destinado, por Japón, a esta misma materia.

Por otro lado, autores como Kim et al. (2011) [12] han comparado las tendencias de suicidios existentes en ambos países; encontrando que, el pico de suicidios más destacado se produce, en Japón, en el grupo de varones con edades comprendidas entre los 45 y los 64 años; mientras que, en Corea del Sur, el pico más importante referente a las tasas de mortalidad por suicidio, es alcanzado en el grupo de varones con edades superiores a los 64 años. En este aspecto, Kino et al. (2019) [3] aplicaron un modelo edad- periodo- cohorte, a las tasas de mortalidad por suicidio de Japón y Corea del Sur comprendidas entre 1986 y 2015, con el objetivo de compararlas. Además, estos autores utilizaron el algoritmo del estimador intrínseco (*Intrinsic Estimator algorithm*) con el objetivo de resolver el problema de identificación asociado a los modelos edad- periodo- cohorte; desarrollando, por tanto, un procedimiento análogo al seguido por Yang y Land (2013) [66].

En primer lugar, Kino et al. (2019) [3] hallaron un importante efecto, en las tasas de mortalidad por suicidio de ambos países, asociado a la variable edad. En este aspecto, dichos autores encontraron un notable efecto de la variable edad en coreanos mayores de 60 años; debido, principalmente, a la pobreza asociada a este colectivo (en 2015, un 46 % de los coreanos con 65 o más años, se encontraba en una situación de pobreza). Por otro lado, en relación con Japón, existía un efecto edad que incidía, positivamente, en las tasas de mortalidad por suicidio de los varones de mediana edad; debido, principalmente, a las duras condiciones laborales que existen en el país nipón, como también señalan autores como Kondo y Oh (2010) [67] o Kim et al. (2011) [12].

Por otra parte, Jeon y Reither (2016) [68] y Kino et al. (2019) [3] encontraron un efecto periodo relacionado con las tasas de mortalidad por suicidio de ambos países. En este aspecto, respecto al país nipón, dichos autores encontraron un importante efecto del regresor periodo, en el segmento de varones, entre 1997 y 2003; coincidiendo, esta etapa, con el periodo conocido como "*The lost 20 Years*" (etapas posteriores al estallido de la crisis asiática de 1997- 1998; caracterizándose, estas, por una etapa de estanflación). Durante dicha época, se produjo un aumento del número de despidos entre las generaciones más jóvenes; además de, un incremento de la precariedad laboral, incidiendo estos hechos, pues, en el aumento de las tasas de mortalidad por suicidio japonesas. Por otro lado, en relación con Corea del Sur, el efecto asociado al regresor periodo se caracterizaba por ser superior al presentado por Japón. En este aspecto, los picos máximos asociados a las tasas de mortalidad por suicidio se corresponden, desde un punto de vista temporal, con los periodos de crisis que ha atravesado el país (1997- 1998, 2003- 2005 y 2008- 2009). De tal forma, las tasas de mortalidad por suicidio coreanas presentaban una cuantía elevada, independientemente del grupo de edad y del sexo; aunque las personas con edades avanzadas, presentaban las mayores ratios de mortalidad por suicidio. Por otra parte, como señalan Kino et al. (2019) [3], las disimilitudes encontradas en los efectos de la variable periodo, entre ambos países asiáticos, pueden estar motivadas por las diferencias que existen, entre ambos países, en materia de políticas sociales. En este aspecto, el sistema público de pensiones japonés se caracteriza por ser generoso con los ancianos; mientras que, en Corea del Sur, las prestaciones a personas de avanzada edad se encuentran infra financiadas.

Por otro lado, en relación con la variable cohorte, Kino et al. (2019) [3] encontraron que, las cohortes de nacimiento japonesas más antiguas; además de, las más recientes, poseían tasas de mortalidad por suicidio más elevadas que la cohorte de referencia (1951); debido a que, las cohortes de nacimiento pertenecientes al *Baby Boom* japonés (personas nacidas en torno a 1951), poseían mejores condiciones laborales que las cohortes de nacimiento más actuales. Estos hechos también han sido contrastados por Brinton (2010) [69]. Por otra parte, también existe un efecto de cohorte importante asociado a las cohortes nacidas antes de 1930; siendo, dichas generaciones, las que sufrieron los efectos más duros asociados a la 2ª Guerra Mundial; además de, la epidemia de adicción a las drogas que se produjo a raíz de dicho conflicto (Vaughn et al., 1995) [70]. Por otro lado, autores como Marchant et al. (2017) [71] hacen alusión al importante crecimiento que se está produciendo, en Japón, en relación con las tasas de mortalidad por suicidio de los adolescentes; pudiendo estar causado, este hecho, por las altas cifras de pobreza infantil existentes, en el país nipón, en la actualidad (Abe, 2012) [72]. Por otro lado, considerando las tasas de mortalidad por suicidio de Corea del Sur, Kino et al. (2019) [3] no hallaron significación alguna en relación con los efectos asociados a la variable cohorte; a pesar de que, desde un punto de vista económico, ambos países asiáticos han atravesado crisis de naturaleza similar. Esta situación puede estar motivada por las diferencias existentes, entre ambas economías, desde el punto de vista del factor trabajo. En este aspecto, en Corea del Sur, no existe una normativa laboral que proteja a los trabajadores en función de su antigüedad en la empresa; por lo que, las crisis económicas afectan, en forma de despidos o precariedad laboral, tanto a las cohortes más jóvenes como a las que poseen mayor edad. Por tanto, en Corea del Sur, no existen diferencias, en materia de suicidios, motivadas por estas causas. Sin embargo, en Japón, las cohortes de nacimiento del *Baby Boom* poseen mayor protección laboral que las cohortes más jóvenes; por lo que, son en estas cohortes, donde se concentra un importante número de suicidios.

Por otra parte, Skegg y Cox (1991) [2] estudiaron las tasas de mortalidad por suicidio de Nueva Zelanda (periodo 1957- 1986); empleando, para ello, un modelo edad- periodo- cohorte. En este aspecto, dichos autores encontraron evidencias que sostienen que, las cohortes de varones nacidas tras la 2ª Guerra Mundial, poseen ratios de mortalidad por suicidio superiores a los hombres nacidos antes del comienzo de este conflicto. Sin embargo, estos autores no encontraron efectos asociados a la variable cohorte en las tasas de mortalidad por suicidio de mujeres.

En el caso del estudio realizado por Goldney y Katsikitis (1983) [73], únicamente se encontraron efectos referentes al regresor cohorte en las cohortes australianas de varones más jóvenes. Sin embargo, Snowdon y Hunt (2002) [25], hallaron importantes efectos asociados a la variable cohorte en el periodo posterior a la 2ª Guerra Mundial; caracterizándose, este último, por un aumento de las tasas de mortalidad por suicidio en las cohortes nacidas tras la finalización de dicho conflicto. Por otra parte, Goldney y Katsikitis (1983) [73] también hallaron un importante efecto asociado a la variable periodo, en relación con las tasas de mortalidad por suicidio australianas, en el año 1967; fecha en la que se introdujeron, en el país oceánico, una serie de restricciones que limitaban el acceso a los barbitúricos.

Por otro lado, Charlton et al. (1993) [74] y McClure (2000) [75] encontraron que, durante los últimos 50 años, se habían producido cambios importantes, en relación con la edad y el género, respecto a las tasas de mortalidad por suicidio de Inglaterra y Gales; destacando, principalmente, las modificaciones que se habían producido en las ratios de mortalidad por suicidio de varones jóvenes. Por otro lado, Murphy et al. (1986) [38] analizaron las tasas de mortalidad por suicidio, de Inglaterra y Gales, pertenecientes al periodo 1970-1980, no encontrando efectos significativos asociados a la variable cohorte. Sin embargo, Surtee y Duffy (1989) [86] y Charlton et al. (1993) [74] encontraron efectos significativos asociados a este último regresor, especialmente en las cohortes más jóvenes. Por otra parte, Amos et al. (2001) [77] hallaron evidencias que señalan que, las medidas legislativas relacionadas con el control de ciertos métodos de suicidios (cambio de gas de carbón a gas natural en los domicilios, mayor control en el acceso a determinados medicamentos, introducción de catalizadores en los motores de combustión interna de los automóviles, ...), habían provocado una modificación de los patrones de suicidios en Inglaterra y Gales. Por otro lado, Gunnell et al. (2003) [78] analizaron las tasas de mortalidad por suicidio, de Inglaterra y Gales, pertenecientes al periodo 1950-1999; utilizando, para ello, métodos gráficos. De tal forma, según autores como Allebeck et al. (1996) [79], la utilización de métodos analíticos, en relación con la aplicación de modelos edad- periodo- cohorte, no

aporta importantes beneficios, en comparación con el método gráfico; por lo que, en base al principio de parsimonia, no tendría sentido avanzar más allá de dicha metodología visual. En este aspecto, en relación con los resultados obtenidos por Gunnell et al. (2003) [78], se observaba un efecto de la variable edad en las tasas de mortalidad por suicidio de mujeres; caracterizándose, este último, por un incremento de las ratios de mortalidad por suicidio a medida que aumentaba la edad. Por otra parte, considerando el segmento de mujeres, se observaba un efecto del regresor periodo en los años 1965- 1970 y finales de los años 70- principios de la década de los 80; caracterizándose, respectivamente, por una reducción y un incremento de las tasas de mortalidad por suicidio. En este aspecto, el descenso de las tasas de mortalidad por suicidio acaecido a finales de los años sesenta del siglo XX; estaba motivado, según autores como Kreitman (1976) [80], por el comienzo de la utilización, a nivel doméstico, del gas natural, en detrimento del gas extraído del carbón. Por otro lado, el incremento de las ratios de mortalidad por suicidio acaecido a finales de los años setenta, estaba asociado con un incremento del desempleo; que se encontraba causado, a su vez, por recesiones económicas, como señalan autores como Gunnell et al. (1999) [81]. Por otra parte, también se observaba un efecto del regresor cohorte; ya que, el valor máximo de las tasas de mortalidad por suicidio de varones se alcanzaba, a medida que aumentaba el tiempo, a edades más tempranas. En otras palabras, el máximo asociado a las tasas de mortalidad por suicidio de varones se alcanzaba, para las cohortes de nacimiento más recientes, a edades más tempranas. En este aspecto, autores como Christophersen et al. (1998) [82], Agid et al. (1999) [83], Platt (2000) [84] y Gunnell et al. (2003) [78] asociaban estos efectos observados en la variable cohorte a:

- El crecimiento de las adicciones relacionadas con ciertas sustancias psicotrópicas.
- El incremento de los problemas matrimoniales (circunstancia que incidía, negativamente, en el desarrollo de los niños).
- Los cambios surgidos en el mercado laboral británico.
- La proliferación de mensajes en los medios de comunicación que mostraban estilos de vida idílicos, generando frustración entre los individuos (pudiendo derivar, incluso, en depresiones que acabasen en suicidios).

Además, la preferencia cada vez mayor de los varones por métodos de suicidio con mayores tasa de letalidad, como el ahorcamiento (en detrimento, por ejemplo, del envenenamiento), pueden explicar estos efectos asociados al regresor cohorte, como señalan autores como Pounder (1993) [85]. Sin embargo, estas conclusiones difieren de las observadas en las tasas de mortalidad por suicidio de mujeres; no observándose, en este caso, efectos asociados al regresor cohorte. Sin embargo, Surtees y Duffy (1989) [86] hallaron que, las cohortes de nacimiento más recientes de mujeres inglesas y galesas, presentaban tasas de mortalidad por suicidio decrecientes.

Una vez realizada la correspondiente descripción de la literatura relacionada con el análisis de las tasas de mortalidad por suicidio, se procederá a describir la base de datos empleada en este estudio.

2.3. Descripción de la base de datos

En primer lugar, tanto las cifras brutas de fallecidos por suicidio como, las tasas de mortalidad por suicidio, proceden de los datos publicados por la Oficina de las Estadísticas Nacionales británica (*Office for National Statistics*)⁹. En este aspecto, las tasas de mortalidad por suicidio (por 100000 habitantes); así como, las cifras brutas de suicidios, han sido calculadas mediante intervalos de periodos y edades de 5 años; segmentando, estos, en función del sexo de la víctima. De tal forma, los 7 periodos que se extienden desde 1985- 1989 hasta 2015- 2019 han sido agrupados, conjuntamente, en una tabla que cubre 15 grupos de edad (15- 19 años hasta 85- 89 años); reflejando, la diagonal de dicha tabla, 20 cohortes de nacimiento que comienzan en 1900- 1904. Por tanto, los datos que se emplearán en el análisis, se encuentran contenidos en un *age- period array*.

⁹Se puede acceder a esta base de datos a partir de la siguiente dirección:
<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/suicidesintheunitedkingdomreferencetables>

Por otro lado, aunque la base de datos publicada por *Office for National Statistics* posee datos de los grupos de edades comprendidos entre los 10- 14 años y los 90 o más años, dichos intervalos no han sido incluidos en el estudio. En este aspecto, en relación con el grupo de personas que poseen entre 10 y 14 años, las tasas de mortalidad por suicidio se caracterizan por ser valores estimados; ya que, el número de suicidios que se produce en esta franja de edad es escaso, en relación con el total poblacional correspondiente a dicho intervalo. Por otro lado, el grupo de edad que representa a la población con 90 o más años, ha sido excluido del estudio; ya que, dicho grupo, refleja un intervalo abierto; siendo, especialmente complicado, determinar la cohorte de nacimiento asociada a las personas que se encuentran en dicho intervalo. En este aspecto, únicamente un 0,747% (N= 1330) de los suicidios acaecidos en el periodo 1985- 2019 fueron realizados por personas con 90 o más años. En relación con varones, dicho porcentaje es del 0,566% (N= 747); mientras que, en mujeres, es del 1,264% (N=583). Por tanto, la decisión de incluir a este grupo de edad generaría unos costes superiores (en términos de inexactitud a la hora de obtener la cohorte de nacimiento) a los beneficios que se obtendrían con la inclusión de dicho intervalo.

Por otra parte, la base de datos publicada por *Office for National Statistics*, poseía las edades de los individuos agrupadas en intervalos de 5 años (desde el grupo de 10- 14 años hasta el intervalo de 90 o más años); mientras que, la variable periodo, poseía una naturaleza de medición anual. De tal forma, ante esta situación, se pueden adoptar dos posibles procedimientos, mutuamente excluyentes, a la hora de estimar un modelo edad- periodo- cohorte:

- Agrupar los distintos periodos en intervalos que posean la misma amplitud que la variable edad. Para ello, se deben realizar medias y, en algunas ocasiones, prescindir de determinados periodos. Por tanto, a través de esta metodología, se pierde información.
- Construir modelos que posibiliten la introducción de intervalos con distintas amplitudes; en este caso, respecto al regresor periodo. Esta metodología ha sido implementada por autores como Holford (2006) [88].

En este caso de estudio, como se señaló anteriormente, se ha utilizado la primera de estas metodologías; agrupando en intervalos de 5 años, pues, a la variable periodo.

Una vez explicadas estas cuestiones, se procederá a introducir los datos utilizados en el estudio. En primer lugar, en el cuadro 2.1, se introducen los suicidios totales en función de los grupos de edad y los intervalos de periodos descritos anteriormente. En este aspecto, las primeras 16 filas de la tabla hacen referencia a las muertes totales (no segmentando por sexo); mientras que, las siguientes 16 filas, hacen referencia a los suicidios de varones. Finalmente, las filas restantes, se corresponden con los suicidios realizados por mujeres.

De tal forma, en relación con los datos contenidos en el cuadro 2.1, el número total de suicidios acaecidos entre 1985 y 2019 es de 176746. Si se desagrega esta cifra por sexos, el 74,232% de las personas que se suicidaron, en dicho periodo, eran varones; mientras que, el 25,768% eran mujeres. Por tanto, y en base a estas notables diferencias, se realizará un análisis tanto a nivel agregado como segmentando (en función del sexo del fallecido).

Por otro lado, en el cuadro 2.2, se introducen las tasas de mortalidad por suicidio, por 100000 habitantes, en función de los grupos de edad y los intervalos de periodos descritos anteriormente. De tal forma, las primeras 16 filas de dicha tabla, hacen alusión a las tasas de mortalidad por suicidio totales (no segmentando por sexo); mientras que, las siguientes 16 filas, hacen referencia a las tasas de mortalidad por suicidio de varones. Finalmente, las filas restantes (últimas 16 filas del cuadro), se corresponden con las tasas de mortalidad por suicidio de mujeres.

Por tanto, en relación con los datos contenidos en el cuadro 2.2, la tasa de mortalidad por suicidio, considerando el periodo 1985- 2019, es del 11,61%. En este aspecto, la tasa de mortalidad por suicidio de varones, para el citado periodo, es del 18,05%; mientras que, en el caso de las mujeres, esta cifra es del 6,08%. Por tanto, las diferencias existentes en las tasas de mortalidad por suicidio, entre ambos grupos de edad, son evidentes; siendo necesario, pues, realizar un análisis desagregado en función del sexo.

Cuadro 2.1: Suicide mortality for periods and age groups (5- year intervals), according to sex, England and Wales, 1985 to 2019. Source: *Office for National Statistics*.

Age/ Period	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019	Totals
15-19	1009	847	852	803	656	667	901	5735
20-24	2362	2606	2096	1701	1467	1610	1777	13619
25-29	2290	2909	2883	2236	1787	1767	2093	15965
30-34	2328	2655	3138	2778	2100	1989	2286	17274
35-39	2438	2471	2656	2962	2706	2347	2238	17818
40-44	2425	2603	2472	2572	2782	2993	2557	18404
45-49	2196	2432	2353	2294	2442	3044	3085	17846
50-54	2138	1915	1979	2122	2081	2712	2910	15857
55-59	2157	1641	1529	1779	1883	2033	2279	13301
60-64	2157	1454	1189	1142	1379	1561	1640	10522
65-69	1961	1420	1065	985	923	1080	1234	8668
70-74	1703	1214	992	904	733	839	988	7373
75-79	1525	1033	894	829	651	680	735	6347
80-84	1026	856	679	620	629	604	617	5031
85-89	411	474	449	425	371	415	441	2986
Totals	28126	26530	25226	24152	22590	24341	25781	176746
15-19	767	674	641	609	482	499	627	4299
20-24	1920	2174	1731	1354	1195	1281	1372	11027
25-29	1812	2386	2365	1773	1394	1408	1623	12761
30-34	1797	2094	2578	2249	1673	1572	1783	13746
35-39	1838	1969	2054	2328	2171	1837	1732	13929
40-44	1797	2019	1882	1983	2217	2402	2006	14306
45-49	1509	1820	1798	1719	1810	2355	2378	13389
50-54	1356	1390	1449	1539	1525	2066	2154	11479
55-59	1408	1152	1091	1261	1374	1541	1731	9558
60-64	1327	957	862	789	974	1179	1204	7292
65-69	1151	917	699	680	641	775	882	5745
70-74	944	713	601	603	527	608	707	4703
75-79	838	593	553	521	449	484	524	3962
80-84	594	504	406	385	415	412	467	3183
85-89	206	272	260	244	245	277	319	1823
Totals (Males)	19264	19634	18970	18037	17092	18696	19509	131202
15-19	242	173	211	194	174	168	274	1436
20-24	442	432	365	347	272	329	405	2592
25-29	478	523	518	463	393	359	470	3204
30-34	531	561	560	529	427	417	503	3528
35-39	600	502	602	634	535	510	506	3889
40-44	628	584	590	589	565	591	551	4098
45-49	687	612	555	575	632	689	707	4457
50-54	782	525	530	583	556	646	756	4378
55-59	749	489	438	518	509	492	548	3743
60-64	830	497	327	353	405	382	436	3230
65-69	810	503	366	305	282	305	352	2923
70-74	759	501	391	301	206	231	281	2670
75-79	687	440	341	308	202	196	211	2385
80-84	432	352	273	235	214	192	150	1848
85-89	205	202	189	181	126	138	122	1163
Totals (Females)	8862	6896	6256	6115	5498	5645	6272	45544

Cuadro 2.2: Suicide mortality rates, per 100,000 inhabitants, for periods and age groups (5- year intervals), according to sex, England and Wales, 1985 to 2019. Source: *Office for National Statistics*.

Age/ Period	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019	Total
15-19	5.36	5.38	5.5	4.88	3.74	3.84	5.46	4.88
20-24	11.44	13.7	13.12	10.54	8.16	8.48	9.52	10.71
25-29	11.98	14.12	14.96	13.06	9.86	9.16	10.44	11.94
30-34	13.72	13.98	15.22	13.98	11.66	10.54	11.54	12.95
35-39	13.82	14.58	14.04	14.3	13.42	12.88	11.76	13.54
40-44	14.46	14.78	14.58	13.56	13.4	14.84	13.98	14.23
45-49	15.62	14.72	13.52	13.64	13	14.76	15.28	14.36
50-54	15.96	13.88	12.08	12.36	12.56	14.5	14.16	13.64
55-59	16.12	12.62	11.34	11.16	11.28	12.48	12.38	12.48
60-64	16.06	11.5	9.54	8.8	9.02	9.68	10.28	10.70
65-69	16.08	11.62	9.14	8.44	7.64	7.38	7.96	9.75
70-74	16.76	11.52	9.34	8.72	6.98	7.54	7.26	9.73
75-79	18.5	12.92	10.5	9.52	7.44	7.46	7.48	10.55
80-84	19.68	15.06	12.08	9.96	9.68	8.84	8.42	11.96
85-89	16.96	16.36	13.8	12.92	9.84	9.92	9.74	12.79
Total Rates	14.83	13.12	11.92	11.06	9.85	10.15	10.38	11.61
15-19	8	8.4	8.12	7.24	5.4	5.6	7.34	7.16
20-24	18.44	22.7	21.56	16.76	13.26	13.3	14.28	17.19
25-29	18.9	23.22	24.6	20.72	15.42	14.62	15.98	19.07
30-34	21.1	22.12	25.18	22.76	18.62	16.7	18.06	20.65
35-39	20.9	23.32	21.84	22.64	21.64	20.22	18.32	21.27
40-44	21.32	23	22.36	21.1	21.54	24.06	22.06	22.21
45-49	21.36	21.96	20.82	20.6	19.36	23.12	23.88	21.59
50-54	20.2	20.14	17.76	18.12	18.5	22.28	21.28	19.75
55-59	21.36	17.82	16.3	15.94	16.68	19.12	19.06	18.04
60-64	20.62	15.6	14.12	12.44	12.96	14.92	15.4	15.15
65-69	20.66	16.12	12.68	12.14	10.98	10.86	11.72	13.59
70-74	21.88	15.58	12.66	12.7	10.72	11.52	10.84	13.70
75-79	26.46	18.88	15.96	14.18	11.68	11.64	11.54	15.76
80-84	35.16	26.22	20.5	16.58	16.2	14.46	14.62	20.53
85-89	33.28	34.74	27.62	24.1	19.5	18.16	18.08	25.07
Total Rates (Males)	21.98	20.65	18.81	17.20	15.50	16.04	16.16	18.05
15-19	2.62	2.26	2.76	2.4	2.02	1.98	3.4	2.49
20-24	4.32	4.58	4.6	4.32	3.02	3.52	4.44	4.11
25-29	5.04	5.06	5.38	5.42	4.36	3.7	4.74	4.81
30-34	6.3	5.88	5.4	5.3	4.76	4.42	5.08	5.31
35-39	6.78	5.92	6.32	6.08	5.28	5.58	5.28	5.89
40-44	7.54	6.62	6.94	6.18	5.4	5.82	5.98	6.35
45-49	9.84	7.42	6.34	6.78	6.68	6.62	6.92	7.23
50-54	11.68	7.58	6.44	6.76	6.66	6.86	7.28	7.61
55-59	10.98	7.48	6.44	6.44	6.02	5.94	5.86	7.02
60-64	11.84	7.62	5.16	5.32	5.24	4.62	5.38	6.45
65-69	12.3	7.68	5.96	5.06	4.5	4.08	4.4	6.28
70-74	12.96	8.4	6.66	5.34	3.7	3.96	3.98	6.43
75-79	13.5	9.02	6.76	6.14	4.14	3.94	4	6.79
80-84	12.3	9.38	7.48	6.02	5.44	4.86	3.64	7.02
85-89	11.42	9.54	8.2	7.96	5.02	5.2	4.42	7.39
Total Rates (Females)	9.29	6.96	6.06	5.70	4.82	4.74	4.99	6.08

Capítulo 3

Fundamentos teóricos de los modelos edad- periodo- cohorte

3.1. Formulación teórica del modelo

Desde un punto de vista estadístico, existe un riesgo importante asociado a confundir los efectos derivados de los regresores edad, periodo y cohorte¹; por lo que, no considerar una de estas variables, a la hora de estimar un modelo de este calibre, podría generar una serie de consecuencias negativas relacionadas con la absorción, por parte de los regresores temporales finalmente introducidos en el modelo, del componente lineal asociado a la variable que ha sido omitida. Sin embargo, si se consideran estas tres variables de forma conjunta, a la hora de realizar el análisis, se generaría el conocido como problema de identificación (Mason et al., 1973) [89]; motivado, este último, por la multicolinealidad derivada de la relación trivial existente entre estas tres variables. De tal forma, considerando el caso de análisis más sencillo, caracterizado por la linealidad existente entre estas tres variables y, empleando el calendario tradicional, se tendría la siguiente expresión:

$$\text{Cohorte} + \text{Edad} = \text{Periodo} + 1 \quad (3.1)$$

En este aspecto, al adoptar esta convención del calendario, no se tendría un hipotético año 0; rodeándose, el factor tiempo, al valor numérico completo más cercano. De tal forma, considerando un ejemplo sencillo en el que los datos fuesen recolectados, anualmente, de forma unitaria se tendría que, el individuo más joven, poseería una Edad = 1; mientras que, la cohorte de nacimiento más reciente, tomaría un valor igual a 1.

Por otro lado, en base a la relación trivial contenida en la expresión 3.1, la pregunta que cabría realizar consistiría en saber si, la relación existente entre la cohorte de nacimiento y la variable dependiente, está supeditada a los efectos de la misma cohorte o a los efectos derivados de la resta entre los regresores edad y periodo.

Por otra parte, estas tres variables pueden ser medidas tanto de forma discreta como de manera continua; aunque, tradicionalmente, se suele emplear la medición de tipología discreta. Normalmente, únicamente dos de los tres regresores temporales son recolectados, generando cierta inexactitud en algunos casos. En este aspecto, cuando se trabaja con los datos asociados a las variables edad y periodo (*age- period array*), se origina el conocido como problema de las cohortes superpuestas (*overlapping cohorts*). Sin embargo, esta situación no afectaría al problema de identificación del modelo, como bien señalan autores como Osmond y Gardner (1989) [90] o Fannon y Nielsen (2018) [91].

¹En este aspecto, la variable edad hace alusión al tiempo acaecido desde el comienzo de un evento (es decir, puede hacer referencia a la edad de un individuo o al tiempo de antigüedad relacionado con una hipotética póliza de seguro). Por otro lado, el regresor periodo hace alusión al punto del calendario en el que se produce el evento (muerte del individuo, reclamación de una póliza de seguro por parte de un cliente, ...). Finalmente, la variable cohorte hace alusión al año en el que nace una persona o se firma, por primera vez, una póliza de seguro entre una determinada compañía y un individuo.

Por otra parte, los datos relacionados con la variable respuesta, pueden ser recolectados a través de diferentes métodos; destacando, en este aspecto, los datos de panel y los datos en formato de corte transversal. De tal forma, desde el punto de vista de la matriz de datos (*data array*), destacan los *AP arrays* (es el que se utilizará en este proyecto), los *AC arrays* y los *PC arrays*. Por otro lado, también son relevantes las matrices de datos en forma de *age-cohort triangle*, que poseen una notable importancia en materia de seguros. En este aspecto, las matrices de datos que se acaban de introducir pueden ser resumidas a través de la siguiente formulación; en la que, se utiliza la notación de trapecoides generalizados propuesta por Kuang et al. (2008) [5]:

$$J = (1 \leq i \leq I \text{ y } 1 \leq k \leq K \text{ y } L + 1 \leq j \leq L + J) \quad (3.2)$$

En este aspecto, i representa la variable edad, k refleja la cohorte y j representa la variable periodo; por lo que, $i + k = j + 1$, en base a la expresión 3.1. Por otra parte, el parámetro L se denomina *periodo de compensación*. Además, I , K y J hacen alusión al número de índices correspondientes a las variables edad, cohorte y periodo, respectivamente. Por tanto, el último periodo de la matriz de datos, tendría un índice igual a $L + 1$. De tal forma, considerando una matriz de datos edad- periodo (*age- period array*) se tendría que, al comienzo de la relación, $L = I - 1$; siendo $L + J = K$. En este aspecto, los modelos más utilizados en el procedimiento de estimación son: los mínimos cuadrados, de tipo Poisson y logísticos, debido a la naturaleza asociada a las matrices de datos que se acaban de introducir².

Por otro lado, los modelos edad- periodo- cohorte poseen un rol destacado en instrumentos como las *life tables* (Bell y Miller, 2005) [92], los *diagramas de Lexis* (Stockard y O'Brien, 2002) [93] o los *Frost's Paper* (Comstock, 2001) [94]. Además, estos modelos poseen un rol destacado en materias de índole económica, demográfica o epidemiológica, por citar únicamente algunos ejemplos³.

Por otra parte, los modelos edad- periodo- cohorte pueden ser vistos como extensiones de un modelo lineal generalizado (GLM). Por otro lado, dichos modelos pueden presentarse, desde un punto de vista algebraico, de distintas maneras; aunque, generalmente, estos son descritos a través de un predictor, de carácter lineal, que posee la siguiente forma (se supondrá un *age- period array*, en base a los datos con los que se trabajará en la parte analítica del estudio- sección 5):

$$\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta \quad (3.3)$$

donde μ_{ij} es el predictor del modelo para la cohorte i y el periodo j , α_i es el efecto de la variable edad del i -ésimo grupo de edad, β_j se corresponde con el efecto del periodo asociado al j -ésimo periodo y γ_k es el efecto de la k -ésima cohorte. En este caso, $k = I - i + j$, siendo I la cifra de grupos de edad existentes en la base de datos. La expresión del predictor puede ser separada; ya que, posee naturaleza aditiva; siendo, los efectos temporales, funciones asociadas a sus índices correspondientes. Desde un punto de vista vectorial, el predictor, μ , se puede reescribir de la siguiente manera: $\mu = (\mu_{ij})_{ij \in J}$, donde el índice J puede ser obtenido a partir de la ecuación 3.2. Dicho predictor es n - dimensional; por lo que, $n = AP$ en el caso de un *age- period array*.

Por otro lado, partiendo de la ecuación del predictor lineal que se introdujo anteriormente, se tiene la siguiente expresión, en base a O'Brien (2014; pp. 24-25)⁴[104], suponiendo una base de datos en forma de *age- period array* rectangular:

$$Y_{ij} = \lambda + \mu_{ij} + \epsilon_{ij} \quad (3.4)$$

En este aspecto, Y_{ij} se corresponde con el valor que toma la variable dependiente para la ij -ésima celda de la base de datos, λ hace alusión a la constante del modelo y ϵ_{ij} es el error correspondiente a la ij -ésima celda de la base de datos. Además, μ_{ij} se corresponde con el predictor lineal; que se encuentra contenido, a su vez, en la ecuación 3.3.

²Dichos modelos se incluyen en la familia de modelos lineales generalizados (GLMs).

³En Fannon y Nielsen (2018) [91] se puede consultar un importante número de aplicaciones relacionadas con los modelos edad- periodo- cohorte.

⁴Esta notación ha sido modificada ligeramente, con el objetivo de garantizar su adecuación a los objetivos del estudio

Por otro lado, en algunos casos se observan, únicamente, las respuestas (*responses*), Y_{ij} ; por lo que, las dosis (*doses*) o exposiciones (*exposures*), Z_{ij} , son desconocidas. En el caso de que estas sean conocidas, es conveniente emplear un modelo dosis- respuestas de tipo Poisson (*Poisson doses- response model*), como se aplicará en este proyecto al analizar las cifras brutas de suicidios de Inglaterra y Gales. De tal forma, al emplear un modelo dosis- respuesta de tipo Poisson, el logaritmo de la expectativa condicional de la distribución de las respuestas (teniendo en cuenta las dosis o la exposición) adquiere una disposición de tipología edad- periodo- cohorte. Equivalentemente, si se diera una modalidad de edad- periodo- cohorte a la expectativa del logaritmo de las respuestas (sin considerar dosis o exposiciones), se obtendría un modelo de respuesta de tipo Poisson. Estos modelos han sido utilizados, mayoritariamente, en estudios epidemiológicos (Martínez- Miranda et al., 2014) [87] o de seguros (*chain- ladder model*).

Por otra parte, también se podría utilizar un modelo de tipo dosis- respuesta logístico. En este caso, las respuestas, Y_{ik} se distribuirían, de forma binomial, considerando las dosis, Z_{ik} , y en base tanto a un parámetro de recuento como a la mortalidad (es decir, la probabilidad de muerte, en este caso), π_{ik} . Por tanto, al considerar un modelo de respuesta logístico, el predictor poseería la siguiente estructura

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \mu_{ij} \quad (3.5)$$

Por otro lado, el principal problema asociado a la utilización de los modelos edad- periodo- cohorte consiste en la imposibilidad de estimar, a través de la implementación de un modelo de estimación estándar, los efectos de cada uno de estos componentes lineales de forma única y separada (al ser términos linealmente dependientes que se encuentran sujetos, por tanto, a multicolinealidad), generando el conocido como problema de identificación del modelo. Esta situación asociada a la parte lineal de los modelos edad- periodo- cohorte, será explicada en el siguiente epígrafe.

3.2. El problema de identificación

Como se acaba de señalar, existe un problema de identificación asociado a los elementos que forman el predictor (edad, periodo y cohorte); ya que, aunque se modificasen los valores de estas variables independientes, se podrían obtener predictores con el mismo resultado (debido a la relación lineal subyacente a los regresores del modelo). En este aspecto, los efectos temporales asociados a estos regresores podrían descomponerse en una componente lineal (sujeta al problema de identificación) y en una parte no lineal (estimable de forma única).

De tal forma, la ecuación del predictor introducida en la expresión 3.3 puede ser reescrita de la siguiente forma, considerando un modelo lineal simplificado:

$$\mu_{ij} = (\alpha_i + a + i \times d) + (\beta_j + b - j \times d) + (\gamma_k + c + k \times d) + (\delta - a - b - c - I \times d) \quad (3.6)$$

siendo a , b , c y d números reales que pueden tomar valores arbitrarios. Por tanto, los efectos temporales asociados a las variables edad, periodo y cohorte únicamente pueden estimarse, desde un punto de vista lineal, considerando una tendencia aleatoria (existen pues, en este plano, infinitas soluciones igualmente válidas). Por tanto, no existe método alguno que permita estimar, de forma única, la tendencia lineal asociada a los datos⁵. Para explicar esta característica, se pueden definir los efectos temporales del modelo a partir del siguiente vector:

$$\theta = (\alpha_1, \dots, \alpha_I; \beta_1, \dots, \beta_J; \gamma_1, \dots, \gamma_K; \delta)' \in \mathbb{R}^q \quad (3.7)$$

donde $q = I + J + K + 1 = 2(I + J)$; siendo, además, μ , la base de datos de edad- periodo (*age- period array*) con la que se trabajará. Dichos datos están asociados, pues, al predictor μ_{ij} . Por tanto, como se señala en Martínez- Miranda et al. (2014) [87], se podrían obtener dos efectos temporales distintos, $\theta^\dagger \neq \theta^\ddagger$, que poseyeran asociado el mismo predictor, $\mu^\dagger = \mu^\ddagger$; ilustrando, este hecho, el problema de identificación

⁵Sin embargo, se podría analizar la parte no lineal referente a los regresores del modelo; ya que, esta última, se expresa a partir de derivadas de, al menos, segundo orden.

asociado a la parte lineal del modelo. Por tanto, el vector que representa al predictor, μ_{ij} , se desplaza por una superficie de, como mínimo, dimensión q . En este aspecto, al existir multicolinealidad en relación con los efectos temporales del modelo, no todos los elementos del vector θ , de dimensión q , pueden obtenerse; por lo que, como el modelo se encuentra sobre- parametrizado, se debe realizar una tarea de parametrización. De tal forma, a pesar de que, los efectos lineales del modelo no pueden ser identificados, se podrían realizar inferencias sobre el predictor, μ_{ij} ; tales como, predecir los valores del mismo (esta característica de los modelos edad- periodo- cohorte será explicada en posteriores epígrafes).

Por otro lado, este problema también puede expresarse desde un punto de vista matricial; siendo, la formulación matemática del mismo, la siguiente:

$$y = Xb + \epsilon \quad (3.8)$$

De tal forma, X es la conocida como matriz de diseño; agrupando, esta, la constante y cada uno de los grupos de edad, periodo y cohorte; y excluyendo, a su vez, los componentes utilizados como elementos de referencia. En este aspecto, las columnas de la matriz de diseño estarían formadas por:

- La constante.
- Los elementos correspondientes a los grupos de edad desde α_1 hasta α_{I-1} .
- Los elementos correspondientes a los periodos situados entre β_1 y β_{J-1} .
- Los elementos que hacen referencia a las cohortes; partiendo desde γ_1 hasta γ_{I+J-2} .

De esta manera, I hace alusión al número de grupos existentes en relación con la variable edad, J hace referencia al número de periodos e $I + J - 1$ se corresponde con la cifra de cohortes existentes. Por tanto, el número de filas de la matriz de diseño, X , es $I \times J$; mientras que, el número de columnas de dicha matriz, es igual a $m = 2(I + J) - 3 = 1 + (I - 1) + (J - 1) + (I + J - 2)$. La categorización de la edad, periodo y cohorte puede realizarse de diferentes maneras; destacando, por ejemplo, la codificación mediante variables *dummies* o la compilación en función de los efectos. Por otro lado, el vector de salidas, y , posee un número de filas igual a la cifra de celdas existente en la base de datos; mientras que, el vector de solución, b , posee $2(I + J) - 3$ filas. Finalmente, ϵ posee $I \times J$ elementos; siendo, este, el vector de los residuos del modelo.

Por otro lado, si se parte de la ecuación mostrada en 3.8 y se premultiplican ambos lados de la misma por X' , se obtiene lo siguiente:

$$X'y = X'(Xb + \epsilon)$$

$$X'y = X'Xb + X'\epsilon$$

Asumiendo que $E(X'\epsilon) = 0$, la anterior ecuación queda de la siguiente forma:

$$X'y = X'Xb \quad (3.9)$$

Para resolver esta última ecuación y obtener el vector b , se premultiplican ambos lados de la expresión por $(X'X)^{-1}$; obteniéndose, a través de este método, el siguiente resultado ⁶:

$$b = (X'X)^{-1}X'y \quad (3.10)$$

Si la matriz de diseño posee un rango completo de columnas, la ecuación tendría solución única. Además, asumiendo que el modelo edad- periodo- cohorte presenta una especificación correcta, se obtendrían estimaciones de los parámetros insesgados (considerando, igualmente, que X posee un rango completo de columnas, hecho que no ocurre en la práctica). De tal forma, como $y = Xb + \epsilon$ según la ecuación 3.8, siendo $b = (X'X)^{-1}X'y$, se obtendría lo siguiente:

⁶Sin embargo, este procedimiento no puede realizarse, en la práctica, a la hora de estimar un modelo edad- periodo- cohorte; ya que, X no posee un rango completo de columnas

$$\begin{aligned}
b &= (X'X)^{-1}X'y; \\
b &= (X'X)^{-1}X'(Xb + \epsilon) \\
b &= b + (X'X)^{-1}X'\epsilon
\end{aligned}$$

Por tanto, como $E[X'\epsilon] = 0$, se obtiene que $E[b] = b$; siendo, pues, insesgado en el caso en el que, la matriz de diseño, X , posea rango de columnas completo.

Sin embargo, en los modelos edad- periodo- cohorte existe una dependencia lineal entre las variables temporales, como se señaló anteriormente. Por tanto, en este caso, el producto de matrices regulares $(X'X)^{-1}$ no podría ser calculado, al no existir. De hecho, no se pueden utilizar métodos de resolución de ecuaciones tradicionales para resolver la expresión 3.10. Con el objetivo de superar este problema, algunos paquetes informáticos eliminan una de las variables independientes (asumiendo que su efecto es igual a 0); obteniendo, posteriormente, una solución de la ecuación 3.9 mediante la implementación del método de mínimos cuadrados. Por tanto, tras eliminar uno de los regresores, se podría calcular la expresión $(X'X)^{-1}$, al no existir dependencia lineal entre las variables independientes. Es decir, al imponer una restricción sobre el modelo, se podría resolver el problema de identificación asociado al mismo; aunque, existiría un número infinito de soluciones. De tal forma, en este punto, resulta fundamental considerar que preguntas pueden ser contestadas en relación con los modelos edad- periodo- cohorte. En este aspecto, podría estimarse, por ejemplo, el predictor μ_{ij} . Además, se podrían obtener conclusiones relacionados con las partes no lineales del modelo. Sin embargo, cuestiones relacionadas con las pendientes o los niveles de los efectos temporales, no pueden ser obtenidas de forma individual, debido al problema de identificación asociado a la parte lineal del modelo⁷.

De tal forma, el problema de identificación del modelo puede ser evitado, siempre y cuando la investigación no esté relacionada con el plano lineal del modelo (McKenzie, 2006) [95]. Sin embargo, en muchas ocasiones, es necesario considerar la parte lineal del modelo; por lo que, se debe considerar el problema de identificación asociado a dicha estructura. De tal forma, para tratar de solucionar este problema; se han utilizado, tradicionalmente, diferentes procedimientos:

- Reparametrizar el modelo edad- periodo- cohorte, de forma que se puedan obtener los efectos temporales asociados a los elementos no lineales de los regresores; siendo, estos, invariantes (Kuang et al., 2008) [5]. Por tanto, a través de esta metodología, se dividirían los efectos temporales del modelo en una parte lineal (sujeta al problema de identificación) y en una parte de índole no lineal (invariante ante cualquier cambio y, por tanto, estimable de forma única). Esta metodología se conoce como *parametrización canónica*.
- Incorporación de restricciones en la parte lineal del modelo, con el objetivo de obtener los efectos temporales asociados al mismo. Estas restricciones deben estar fundamentadas en la teoría; aunque, la veracidad de las mismas no puede ser demostrada desde un punto de vista analítico.
- Recurrir a la metodología de cálculo de las funciones estimables; ya que, dichas funciones son invariantes ante las restricciones tomadas. De tal forma, a través de estas funciones, pueden ser obtenidos diversos elementos de interés.
- Utilizar métodos bayesianos con el objetivo de solventar el problema de identificación. Como se verá posteriormente, esta metodología genera una serie de problemas similares a los que se obtienen al resolver el modelo edad- periodo- cohorte mediante la imposición de restricciones.
- Abandonar la estructura de los modelos edad- periodo- cohorte. En este aspecto, se pueden sustituir una, o varias, de las variables temporales del modelo por otros regresores ajenos a estas (aproximación factor- característica). Esta metodología ha sido seguida por autores como Attanasio (1998) [96] o Browning et al. (2016) [97]. Por otro lado, también se podría seguir una metodología de análisis gráfica; siendo, esta última, ampliamente utilizada en la práctica.

Estos procedimientos de resolución serán descritos en los siguientes epígrafes.

⁷En el epígrafe 3.6 se hará una revisión de las cuestiones que pueden ser resueltas mediante la utilización de los modelos edad- periodo- cohorte, considerando el problema de identificación asociado a estos.

3.3. Procedimientos para abordar el problema de identificación

Como se señaló anteriormente, existe un conjunto importante de métodos que pretenden resolver el problema de identificación inherente a los modelos edad- periodo- cohorte. En este epígrafe, se realizará una descripción de estos métodos; aunque, en primer lugar, se hará un resumen de las estrategias seguidas para abordar el problema de identificación.

De tal forma, en primer lugar, es fundamental trabajar con funciones temporales que no estén sometidas a las variaciones expresadas en la ecuación 3.6. En este aspecto, autores como Carstensen (2007) [98], consideran la expresión 3.6 como un conjunto g que transforma los efectos temporales contenidos en θ (ver ecuación 3.7) y los convierte en el vector de predictores $\mu = (\mu_{ij})_{ij} \in J$. Por tanto, $f(\theta)$ es invariante si $f\{g(\theta)\} = f(\theta)$; siendo, por tanto, $f(\theta)$ una función invariante de θ , en dicho caso. Considerando esta definición, las dobles diferencias asociadas a los efectos temporales presentarían la propiedad de invarianza, como se mostrará a través del siguiente ejemplo, en el que se consideran las dobles diferencias referentes a los efectos de la variable edad ⁸:

$$\Delta^2 \alpha_i = \Delta \alpha_i - \Delta \alpha_{i-1} = \alpha_i - 2\alpha_{i-1} + \alpha_{i-2} \quad (3.11)$$

En este caso, las dobles diferencias de la variable edad serían invariantes a las transformaciones asociadas a la expresión 3.6; ya que, las dobles diferencias referentes a los efectos temporales del regresor edad, no dependerían de los parámetros a y d .

Por otro lado, el predictor μ_{ij} también es invariante ante cualquier transformación, como señalan autores como Kuang et al. (2008) [5]. Sin embargo, el predictor, per se, no reviste de un especial interés; siendo, lo realmente relevante, la representación del predictor expresado en base a funciones invariantes, $\xi(\theta)$. Por tanto, el interés consiste en expresar el predictor, μ , en función de $\xi(\theta)$; siendo, ξ invariante; por lo que, $\xi(\theta) = \xi(g(\theta))$. En este aspecto, ξ es una función invariante (de acuerdo al problema de identificación del modelo), siendo especialmente utilizada a la hora de parametrizar el modelo. Además, si ξ es una función lineal asociada a los efectos temporales del modelo; variando, también, es un espacio paramétrico no cerrado, ξ sería un parámetro canónico.

Por otra parte, también se podría considerar la propiedad de invarianza a partir del análisis de submuestras. En este aspecto, dicho procedimiento permite analizar estrategias de identificación que pretenden evitar el problema de identificación asociado a los modelos edad- periodo- cohorte. Además, el análisis de submuestras puede utilizarse a la hora de realizar pruebas de especificación.

De tal forma, para que un determinado método de estimación evite el problema de identificación, los resultados de este deberán ser invariantes, independientemente de la submuestra empleada. Por ejemplo, si se consideran dos base de datos, J y J' (ver expresión 3.1); estando formadas, respectivamente, por la base de datos completa y una submuestra de la misma (representativa de la base de datos original), un método que no está influido por el problema de identificación arrojará, independientemente de la muestra empleada (J o J' , en este caso), los mismos resultados. En este aspecto, se podría implementar un análisis recursivo, con el objetivo de estudiar la especificación subyacente al modelo estimado. Para ello, se podrían seguir los siguientes pasos, como se señala en Chow (1960) [99]:

1. Escoger distintas submuestras (J'), en función de una serie de valores umbrales diferentes.
2. Estimar los parámetros asociados a dichas submuestras y representarlos, posteriormente, en función de los valores umbrales elegidos en el paso anterior.
3. Estudiar la especificación asociada al modelo mediante un análisis recursivo; considerando, para ello, los efectos temporales referentes a las variables edad, periodo y cohorte.

Finalmente, si las variaciones en los resultados, respecto a las diferentes submuestras, son escasas, el modelo se encontrará especificado de forma correcta. Sin embargo, ante variaciones de notable cuantía, se deberá

⁸Este ejemplo ha sido extraído de Fannon y Nielsen (2018) [91]; aunque, ha sido modificado ligeramente, con el objetivo de adecuarlo a la notación matemática utilizada en este estudio

revisar la especificación inherente al modelo.

Una vez expuestas estas cuestiones, se procederá a analizar, en los próximos epígrafes, los diferentes métodos que pueden utilizarse a la hora de abordar el problema de identificación inherente a la parte lineal de los modelos edad- periodo- cohorte.

3.3.1. Parametrización canónica

Como se señaló anteriormente, los efectos temporales de un modelo edad- periodo- cohorte pueden ser descompuestos en una parte no lineal, que puede representarse a través del sumatorio doble de las dobles diferencias (como se muestra en la ecuación 3.11), y en una parte de carácter lineal. En este aspecto, la constante del modelo, junto con los elementos lineales asociados a los tres efectos temporales de las variables independientes, formarían un *plano lineal*, como se señala en Holford (1983) [100]. Por tanto, en base a estas ideas, el predictor poseería la siguiente expresión:

$$\mu_{ij} = \text{plano lineal} + \sum_i \sum \Delta^2 \alpha_s + \sum_j \sum \Delta^2 \beta_s + \sum_k \sum \Delta^2 \gamma_s \quad (3.12)$$

En este aspecto, los índices de las sumas correspondientes a los dobles sumatorios de las dobles diferencias, unido a la expresión concreta del plano lineal, presentan dependencia directa respecto a los índices de las variables edad, periodo y cohorte. Se debe señalar que, en el caso de la ecuación 3.12, el plano podría identificarse; siendo, además, invariante ante transformaciones como las que se describen en la expresión 3.6. Sin embargo, las pendientes asociadas a los efectos temporales no podrían identificarse; ya que, los regresores edad, periodo y cohorte continuarían poseyendo una relación lineal directa.

Por otro lado, los componentes no lineales de la ecuación 3.12 pueden ser separados del plano lineal; por lo que el predictor, μ_{ij} , podría ser expresado a través de la siguiente expresión:

$$\mu_{ij} = \xi' \times x_{ij} \quad (3.13)$$

donde x_{ij} hace alusión a la matriz de diseño (aunque, en este caso, es el vector de diseño; que se encuentra definido, a su vez, en base a $m(t, s) = \max(t - s + 1, 0)$); mientras que, ξ hace referencia al parámetro de identificación; poseyendo, este último, la siguiente expresión:

$$\xi = (\mu_{I1}, \mu_{I1} - \mu_{I-1,1}, \mu_{I,2} - \mu_{I1}, \Delta^2 \alpha_3, \dots, \Delta^2 \alpha_I, \Delta^2 \beta_3, \dots, \Delta^2 \beta_J, \Delta^2 \gamma_3, \dots, \Delta^2 \gamma_K)' \in \mathbb{R}^p \quad (3.14)$$

siendo $p = q - 4 = 2 \times (I + J - 2)$. Por tanto, el parámetro de identificación está formado por:

- Las dobles diferencias de los regresores.
- Tres componentes del predictor; que fijan, a su vez, la tendencia lineal del mismo.

Además, según el *Teorema 1* contenido en Kuang et al. (2008; pp.981) [5] el parámetro de identificación, ξ (mostrado en la ecuación 3.14), posee las siguientes características:

- ξ es una función de θ (ver expresión 3.7).
- μ es una función de ξ . Se debe recordar que $\mu = (\mu_{ij})_{ij \in J}$.

De tal forma, la parametrización del predictor, μ , según el parámetro de identificación, ξ , puede identificarse; ya que, si $\xi^\dagger \neq \xi^\ddagger$ se tiene que $\mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$. Por tanto, el parámetro de identificación es invariante ante modificaciones como las descritas en la expresión 3.1; estando conformado, además, tanto por las dobles diferencias de los regresores como, por los valores asociados al predictor. En este aspecto, ξ presenta la propiedad de canonicidad, en relación con las familias exponenciales y los modelos normal, logístico y de Poisson.

Por otro lado, como los efectos no lineales del modelo se recogen a través de dobles diferencias, estos poseen una interpretación de diferencias en diferencias (*difference-in-difference*). De igual manera, también

poseen una interpretación de *odds ratio*. Por ejemplo, una doble diferencia referente al regresor edad, poseería la siguiente expresión (considerando una matriz de datos edad- cohorte); siendo equivalente, pues, a la que se muestra en la expresión 3.11:

$$\Delta^2 \alpha_i = \mu_{i,k} - \mu_{i-1,k} - \mu_{i-1,k+1} + \mu_{i-2,k+1} \quad (3.15)$$

donde μ hace referencia al predictor. De igual forma, la convención temporal asociada a la interpretación se encuentra contenida en la expresión 3.1. Además, conviene recordar que, estas segundas diferencias, pueden ser identificadas, debido a las propiedades de las derivadas.

Por otra parte, si se reescribe la expresión 3.15 en términos de un *age- period array* se tiene la siguiente expresión, considerando las dobles diferencias asociadas a la variable edad:

$$\Delta^2 \alpha_i = \mu_{i,j} - \mu_{i-1,j-1} - \mu_{i-1,j} + \mu_{i-2,j-1} \quad (3.16)$$

Respecto a las expresiones 3.15 y 3.16, se podría utilizar cualquier periodo o cohorte, siempre que estos sean elementos consecutivos. Por otro lado, en el caso de variables dependientes binarias, las dobles diferencias poseen una interpretación de tipo *log-odss*, como se señala en Fannon y Nielsen (2018) [91].

Por otra parte, las dobles diferencias presentan la propiedad de invarianza, como se citó anteriormente; aunque, al utilizar este procedimiento, se obtendrán conclusiones algo erráticas. Por tanto, lo que se suele realizar, con el objetivo de facilitar las tareas interpretativas, consiste en acumular las dobles diferencias.

Volviendo a la expresión 3.13, faltaría por encontrar la matriz de diseño que permitiese relacionar al predictor y al parámetro de identificación. De tal forma, la matriz de diseño depende, a su vez, de la base de datos escogida. En este caso, al contar con un *age- period array* en relación con los suicidios; se puede aplicar, con el objetivo de unir al predictor y al parámetro de canonización, el siguiente teorema (Martínez-Miranda et al., 2014) [87]:

Theorem 3.3.1. *Partiendo de una base de datos de tipo edad- periodo, siendo $i = 1, \dots, I$, $j = 1, \dots, J$, $k = I - i - j$, y $\mu_{i,j} = \alpha_i + \beta_j + \gamma_k + \delta$ se tiene lo siguiente:*

$$\mu_{ij} = \mu_{I1} + (i - I)(\mu_{I1} - \mu_{I-1,1}) + (j - 1)(\mu_{I2} - \mu_{I1}) + \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2 \alpha_{s+2} + \sum_{t=3}^j \sum_{s=3}^t \Delta^2 \beta_s + \sum_{t=3}^k \sum_{s=3}^t \Delta^2 \gamma_s$$

Por tanto, en base a esta última expresión, el parámetro de identificación, ξ permite identificar al predictor, μ , de forma que $\xi^\dagger \neq \xi^\ddagger$ provoca que $\mu(\xi^\dagger) \neq \mu(\xi^\ddagger)$.

El teorema 3.3.1 posee una serie de características:

- Permite descomponer el predictor en un nivel general, dos tendencias de carácter lineal y tres efectos temporales.
- No se puede escoger un punto de referencia que permita que, los tres efectos temporales, se incrementen al mismo tiempo; ya que, las diferencias de las variables periodo y cohorte se acumulan hacia adelante; mientras que, el regresor edad, se acumula hacia atrás. En un *age- cohort array*, esta característica sería diferente.
- Por otro lado, las dobles diferencias $\Delta^2 \alpha_i$, $\Delta^2 \beta_j$ y $\Delta^2 \gamma_k$ pueden ser interpretadas en función de *log odds-ratios*; por lo que, pueden reescribirse de la siguiente forma (ver expresión 3.16):

$$\log \left(\frac{\frac{\pi_{ij}}{\pi_{i-1,j}}}{\frac{\pi_{i-1,j-1}}{\pi_{i-2,j-1}}} \right) = \mu_{ij} - \mu_{i-1,j} - \mu_{i-1,j-1} + \mu_{i-2,j-1} = \Delta^2 \alpha_i \quad (3.17)$$

$$\log \left(\frac{\frac{\pi_{ij}}{\pi_{i,j-1}}}{\frac{\pi_{i-1,j-1}}{\pi_{i-1,j-2}}} \right) = \mu_{ij} - \mu_{i,j-1} - \mu_{i-1,j-1} + \mu_{i-1,j-2} = \Delta^2 \beta_j \quad (3.18)$$

$$\log \left(\frac{\frac{\pi_{ij}}{\frac{\pi_{i,j-1}}{\frac{\pi_{i+1,j}}{\pi_{i+1,j-1}}}}}{\frac{\pi_{i,j-1}}{\frac{\pi_{i+1,j}}{\pi_{i+1,j-1}}}} \right) = \mu_{ij} - \mu_{i,j-1} - \mu_{i+1,j} + \mu_{i+1,j-1} = \Delta^2 \gamma_k \quad (3.19)$$

Por otra parte, los parámetros asociados a las pendientes del predictor pueden ser expresados de la siguiente forma:

$$\log \left(\frac{\pi_{I1}}{\pi_{I-1,1}} \right) = \mu_{I1} - \mu_{I-1,1} \quad (3.20)$$

$$\log \left(\frac{\pi_{I2}}{\pi_{I1}} \right) = \mu_{I2} - \mu_{I1} \quad (3.21)$$

3.3.1.1. Interpretación de los efectos temporales

Como los efectos temporales originales de los modelos edad- periodo- cohorte no pueden ser interpretados (debido a la imposibilidad de obtención de los mismos), se deben analizar los efectos temporales asociados a las partes no lineales. Para ello, se debe aplicar el procedimiento conocido como *detrending*, que elimina las tendencias de la serie asociadas al largo plazo, con el objetivo de analizar los cambios que se producen a corto plazo. En este aspecto, para aplicar este método, se deben representar, en tres gráficas distintas, los efectos temporales asociados a las variables edad, periodo y cohorte; por lo que, el análisis debe realizarse de forma individual. Además, las tendencias de carácter lineal únicamente quedarán asociadas al plano lineal.

Por otro lado, no es recomendable utilizar las dobles diferencias acumuladas, desde un punto de vista interpretativo (aunque son útiles en materia de estimación); ya que, la representación gráfica de los efectos temporales no lineales asociados a estas dobles diferencias, presentan una tendencia importante, dificultando la labor interpretativa del investigador. Sin embargo, existe una notable ventaja asociada a este método: la acumulación de las dobles diferencias posee los mismos grados de libertad que las dobles diferencias.

Por otra parte, se puede aplicar el procedimiento de *detrending* a la doble sumatoria de las dobles diferencias. Para ello, los tres conjuntos de dobles sumatorias deben ser ortogonalizados, de forma individual, respecto a una constante y una restricción de carácter temporal. Por tanto, la ortogonalización debe ser aplicada, de manera individual, a las dobles sumas de los regresores temporales, con el objetivo de aislar la tendencia. Sin embargo, no se puede comprobar el hecho de que, las dobles diferencias, posean los mismos grados de libertad que las dobles sumatorias de las dobles diferencias; siendo, este, el principal inconveniente de dicho método.

Por último, también se podría utilizar el procedimiento *detrending*, fijando el comienzo y la conclusión de las dobles sumatorias a cero. De tal forma, en base a Nielsen (2015) [129], esta metodología podría ser implementada de la siguiente manera; considerando, por ejemplo, al regresor edad:

1. En primer lugar, se debe definir el siguiente parámetro:

$$\alpha_i^{\text{detrend}} = \alpha_i^{\sum \sum \Delta \Delta} - a - d \times i$$

2. Posteriormente, partiendo del paso anterior se tiene que $a = -d$, en base a la expresión 3.6. Por tanto:

$$d = \frac{\alpha_A^{\sum \sum \Delta \Delta}}{(A-1)}$$

3. Finalmente, se tiene que las dobles sumatorias comienzan y concluyen en cero:

$$\alpha_1^{\text{detrend}} = \alpha_A^{\text{detrend}} = 0$$

De tal forma, mediante la implementación de este procedimiento, los grados de libertad de las dobles sumatorias serán iguales a los de las dobles diferencias; siendo, esta, una de las principales ventajas asociadas a este procedimiento. Además, debido a las restricciones impuestas en relación con $\alpha_i^{\text{detrend}}$, la no linealidad

puede ser estudiada fácilmente.

Por otro lado, $\alpha_i^{\text{detr}}d$ depende de la matriz de índices. Además, el gráfico asociado a $\alpha_i^{\text{detr}}d$ permite identificar curvas con forma de U o S , cuestión que revista de una notable importancia en términos interpretativos. En este aspecto, dichas curvas pueden ser testadas a través de la aplicación de polinomios cuadráticos o de orden superior.

3.3.2. Identificación del modelo mediante la imposición de restricciones

La metodología más utilizada, desde un punto de vista tradicional, a la hora de enfrentarse al problema de identificación inherente a los modelos edad- periodo- cohorte, consiste en la imposición de restricciones en la parte lineal del modelo. Sin embargo, este método genera una serie de parámetros que no cumplen con la propiedad de invarianza (ver ecuación 3.6). En este aspecto, el objetivo de imponer estas restricciones, consiste en obtener un reflejo de las partes lineales asociadas a los efectos temporales del plano lineal del modelo. De tal forma, no existe una manera única de dividir dichos efectos temporales entre los tres regresores del modelo; ya que, coexisten infinitas soluciones, en el espacio lineal, que dependen de la restricción adoptada.

Por otro lado, en relación con las restricciones que pueden implementarse, existen una serie de aproximaciones teóricas que podrían utilizarse como método de elección. En este aspecto, generalmente, dichos procedimientos metodológicos se desarrollan en dos etapas, como se señala en Nielsen y Nielsen (2014) [101]:

1. Identificación de los niveles asociados al plano lineal.
2. Obtención de la pendiente asociada a dichos niveles.

Estos pasos serán descritos a continuación⁹:

3.3.2.1. Imposición de restricciones sobre los niveles

Desde un punto de vista estadístico, existen dos métodos para identificar los niveles: imponer restricciones en relación con las coordenadas de los efectos temporales o restringir el nivel promedio asociado a los efectos temporales de los regresores. Se debe señalar que, ninguno de estos dos métodos, arroja resultados invariantes. Dichos procedimientos se describirán, brevemente, a continuación:

Imposición de restricciones sobre las coordenadas de los efectos temporales

Este procedimiento ha sido utilizado por autores como Ejrnæs y Hochguertel (2013) [102] y consiste en restringir el valor de los efectos temporales a cero. Sin embargo, estas restricciones tampoco presentan la propiedad de invarianza; aunque, dicho método permite realizar comparaciones entre las diferentes submuestras, siempre y cuando se apliquen las mismas restricciones.

Imposición de restricciones sobre los niveles promedio

Esta metodología ha sido empleada por autores como Schulhofer-Wohl (2018) [103] y consiste en fijar la media de los efectos temporales de los regresores a cero. De tal forma, la restricción que se utilizaría sería la siguiente:

$$\frac{1}{I} \sum_{i=1}^I \alpha_i = \frac{1}{J} \sum_{j=L+1}^{L+J} \beta_j = \frac{1}{K} \sum_{k=1}^K \gamma_k = 0$$

Por tanto, como $\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta$, el nivel del modelo quedará reflejado en base a la constante δ . De esta manera, al imponer esta restricción, se consigue ortogonalizar los niveles y las pendientes de la parte lineal del modelo. Sin embargo, la aplicación de este método, genera una dependencia entre los efectos temporales asociados a los regresores y la dimensión perteneciente a la matriz de índices del modelo; contenida, a su vez,

⁹Se debe señalar que estas aproximaciones han sido extraídas de Fannon y Nielsen (2018) [91].

en la ecuación 3.2. Además, como en los procedimientos anteriores, este método no presenta la propiedad de invarianza; ya que, las transformaciones asociadas a la expresión 3.6, modifican los resultados obtenidos.

3.3.2.2. Imposición de restricciones sobre las pendientes

Tras realizar la correspondiente identificación de los niveles asociados al plano lineal, se deben implementar restricciones que permitan restringir las pendientes. De tal forma, este tipo de restricciones, por definición, se encargan de unir las pendientes correspondientes a los tres efectos temporales, no poseyendo asociada, a su vez, la propiedad de invarianza. En este aspecto, se pueden unir las pendientes de los efectos temporales de las variables edad, periodo y cohorte a partir de los siguientes métodos:

Imposición de restricciones en efectos temporales adyacentes

El objetivo de esta metodología consiste en identificar la pendiente, mediante la incorporación de una restricción basada en la igualación de dos efectos temporales consecutivos. Por ejemplo, se podría establecer una restricción del tipo $\gamma_2 = \gamma_3$; aunque, este tipo de restricciones, se caracteriza por no presentar la propiedad de invarianza.

Imposición de restricciones mediante la ortogonalización de un efecto temporal

Otro de los métodos utilizados a la hora de hacer frente al problema de identificación, consiste en fijar uno de los efectos temporales, con el objetivo de igualar la pendiente de este a cero. En este aspecto, una restricción de este tipo, en relación con la variable periodo, podría adoptar la siguiente expresión:

$$\sum_j j \times \beta_j = 0$$

Sin embargo, al igual que en el resto de casos que se han ido introduciendo, la interpretación variaría en función de la restricción adoptada.

Aplicación de la matriz inversa generalizada

Por otro lado, el problema de identificación referente a los modelos edad- periodo- cohorte, al estar cimentado en el concepto de la colinealidad, podría ser abordado a través de la aplicación de matrices inversas generalizadas (O'Brien, 2011) [105]. A continuación, se presentará el método de construcción de dichas matrices; además de, algunas características asociadas a las mismas. Aprovechando la explicación de esta metodología se explicarán, también, algunas características referentes a los modelos edad- periodo- cohorte.

Para tal fin, se partirá de la ecuación normal 3.10, que poseía la siguiente forma:

$$b = (X'X)^{-1} X'y$$

donde b se corresponde con el vector de soluciones asociado al sistema; poseyendo, este último, infinitas soluciones; ya que, el objeto $(X'X)^{-1}$ no podía ser calculado (al tener rango completo, debido a la multicolinealidad existente entre las variables temporales). En este aspecto, la inversa generalizada presenta la siguiente expresión:

$$(X'X)'$$

Por tanto, aplicando esta matriz a la ecuación normal, se obtiene la siguiente expresión:

$$b_c^0 = (X'X)^- X'y \tag{3.22}$$

donde b_c^0 se corresponde con la solución a la ecuación normal, considerando la restricción asociada a la matriz inversa generalizada $(X'X)^-$. Sin embargo, como se justificó anteriormente, esta no es la única solución asociada a dicha ecuación; ya que, dependiendo del valor de $(X'X)^-$ (que depende, a su vez, de la restricción adoptada) existe un número infinito de soluciones; siendo, cada una de ellas, una solución de mínimos

cuadrados.

De tal forma, para construir la matriz inversa generalizada, se puede seguir el método desarrollado por Mazumdar et al. (1980) [106]:

1. Calcular el producto de matrices $X'X$.
2. Intercambiar la última fila de la matriz $X'X$ (que se acaba de calcular en el paso anterior) por una restricción determinada.
3. Calcular el valor de la inversa de esta última matriz.
4. Partiendo de la matriz inversa que se acaba de calcular, se debe reemplazar la última columna de la misma por una columna formada por ceros (restricción c_1). Tras esto, se obtiene la matriz inversa generalizada, representada a través de la siguiente expresión: $(X'X)_{c_1}^-$. Sustituyendo dicha matriz en la ecuación 3.22, se obtendría la solución $b_{c_1}^0$ asociada a la restricción aplicada (c_1 , en este caso).

Por otro lado, este método permitiría realizar los cálculos en base a la i -ésima fila de la matriz $X'X$. En este aspecto, tras calcular la inversa de dicha matriz, se debería reemplazar la columna i -ésima por una columna de ceros; siguiendo, por tanto, los pasos que se acaban de introducir.

De tal forma, partiendo de este último procedimiento, y considerando que $b_{c_1}^0$ es una de las soluciones asociadas a la ecuación normal, se tiene lo siguiente:

$$X'Xb_{c_1}^0 = X'y \quad (3.23)$$

En este aspecto, la dependencia lineal asociada a los modelos edad- periodo- cohorte está provocada por la existencia de una combinación lineal, en una de las columnas de la matriz X , que produce un vector igual a 0 (impidiendo que se pueda realizar el cálculo $(X'X)^{-1}$). Partiendo de esta idea, se considerará que $Xv = 0$, siendo X la matriz de transición y v el vector nulo. Esta expresión puede ser reescrita de la siguiente forma, donde s es un escalar:

$$Xsv = 0 \quad (3.24)$$

Además, solo existe un vector nulo asociado al modelo edad- periodo- cohorte; ya que, la matriz X , posee un rango deficiente en uno.

Por otra parte, como $Xsv = 0$, $X'Xsv = 0$; se tiene que:

$$X'y = X'Xb_{c_1}^0 + X'Xsv \quad (3.25)$$

que presenta la misma solución que la ecuación 3.23. De tal forma, mediante la agrupación de términos se puede construir la siguiente expresión:

$$X'y = X'X(b_{c_1}^0 + sv) \quad (3.26)$$

Por tanto, $b_{c_1}^0 + sv$ es una solución de la ecuación normal; formando, además, una línea contenida en un espacio multidimensional. De tal forma, la ecuación vectorial, para dicha línea, puede ser expresada a través de la siguiente ecuación:

$$b_c^0 = b_{c_1}^0 + sv \quad (3.27)$$

Cualquier solución de la ecuación normal es un punto de esta línea; que se encuentra, a su vez, en un espacio multidimensional. En este aspecto, v se corresponde con la dirección de dicha línea. Por tanto, cualquier solución que se encuentre sobre esta línea, será una solución de la ecuación normal.

Por otro lado, debido a los problemas asociados a la matriz de diseño, X , únicamente existe un vector de ceros, v , que permita que $Xsv = 0$. Por tanto, solo existe un vector sv que multiplique a $X'X$ y que arroje un valor igual a 0. En base a este hecho, únicamente existe una línea de soluciones relacionada con la

ecuación normal; que posee, a su vez, infinitas soluciones igualmente satisfactorias.

Si se realice el análisis de los modelos edad- periodo- cohorte desde un punto de vista de los modelos lineales generalizados (como el de Poisson, por ejemplo)¹⁰, en detrimento de aplicar el procedimiento de mínimos cuadrados a la hora de resolver la ecuación normal, se obtendrían resultados similares a los que se explicaron anteriormente, en relación con la imposibilidad de encontrar una solución única. Esta característica será demostrada a continuación.

En primer lugar, conviene tener en cuenta que los modelos lineales generalizados pueden expresarse de la siguiente forma:

$$g(E(Y_{ij})) = Xb \quad (3.28)$$

siendo, $E(Y_{ij})$ el valor esperado de la variable dependiente para la ij -ésima celda de la base de datos a analizar; que consistía, a su vez, en una base de datos edad- periodo (age- period array). En este aspecto, g es la función de enlace.

Por otro lado, la matriz de diseño, X , presenta el mismo problema que poseía el modelo estimado por mínimos cuadrados ordinarios; ya que, $(X'X)^{-1}$ no puede calcularse (al existir una relación lineal perfecta entre las variables independientes). Por tanto, se puede considerar la siguiente expresión, al igual que en el caso de la estimación por mínimos cuadrados, $Xv = 0$; por lo que, $Xsv = 0$ siendo s un escalar. Este hecho ilustra, por ende, la dependencia lineal existente entre los regresores edad, periodo y cohorte. Por tanto, existe una línea de soluciones definidas por la expresión $g(E(Y_{ij})) = X(b_{c1}^0 + sv)$ (donde $b_c^0 = b_{c1}^0 + sv$) que proporciona un número infinito de soluciones igualmente válidas.

Por otra parte, en cuanto al ajuste del modelo, los valores predichos de la variable de interés toman el mismo valor, independientemente de la restricción adoptada. En este aspecto, volviendo al modelo general, se había llegado a la conclusión de que todas las soluciones derivadas del ajuste del modelo se situaban sobre la línea de soluciones $b_c^0 = b_{c1}^0 + sv$. Considerando que b_{c1}^0 genera una solución que se encuentra sobre dicha línea, los valores predichos referentes a la variable de interés pueden ser representados a partir de la siguiente expresión:

$$\hat{y} = Xb_{c1}^0 \quad (3.29)$$

Por otro lado, b_c^0 también se encuentra sobre esta línea (al igual que un número infinito de valores); por lo que, produce soluciones tan eficientes como b_{c1}^0 . En este aspecto, como $Xsv = 0$, $Xb_c^0 = Xb_{c1}^0 + Xsv = Xb_{c1}^0 = \hat{y}$; por lo que, independientemente de la restricción adoptada, se obtienen resultados predichos idénticos¹¹. Este hecho imposibilita que, pueda preferirse una determinada restricción, sobre otra, a la hora de estimar el modelo; ya que, independientemente de la restricción adoptada, el ajuste del modelo es el mismo (los valores predichos toman el mismo valor, independientemente de la restricción adoptada).

Por otro lado, la solución asociada al modelo edad- periodo- cohorte es ortogonal a la restricción impuesta para su resolución, como se señala en Mazumdar et al. (1980)[106]. En este aspecto, el producto cruzado de la restricción impuesta y el vector que refleja la solución debe ser igual a 0 (por tanto, ambos vectores son ortogonales). De tal forma, desde un punto de vista matemático:

$$c'b_c^0 = 0 \quad (3.30)$$

reflejando el vector c la restricción.

Por otro lado, las soluciones difieren entre sí en función de sv , al encontrarse sobre una misma línea; siendo s , a su vez, un escalar. Además, los elementos que conforman el vector de ceros se caracterizan por ser

¹⁰Es relevante señalar que, estimando por mínimos cuadrados ordinarios el logaritmo de la variable dependiente, se obtienen resultados similares a los derivados de estimar el modelo considerando la regresión de Poisson (O' Brien, 2003)[107].

¹¹Este resultado también se obtiene, con sus respectivas expresiones, en el caso de la aplicación de modelos lineales generalizados, tales como el modelo Poisson o el Binomial Negativo. Sin embargo, el valor de las estimaciones de la variable dependiente varía en función del método utilizado.

lineales dentro de los propios grupos de edad, periodo y cohorte. De tal forma, el valor de la tendencia, en cuanto al signo de esta en alusión al vector de ceros, no presenta diferencia alguna en relación con los grupos de edad y cohorte; aunque, posee un signo opuesto al periodo (en virtud de la ecuación 3.1). Por tanto, la solución obtenida a través de la adopción de una nueva restricción, desplazará la tendencia de los coeficientes asociados a las variables edad y cohorte en la misma dirección; provocando, a su vez, un desplazamiento de signo contrario en relación con la tendencia de la variable periodo. De tal forma, estos desplazamientos concernientes a las tendencias lineales de los coeficientes asociados a las variables independientes del modelo, originan las diferencias existentes entre las múltiples soluciones; por lo que, las tendencias lineales del modelo, no pueden ser identificadas.

Por otra parte, cuando se dispone de una base de datos clasificada en función de las variables edad y periodo (*age-period array*), una estrategia utilizada tradicionalmente en la literatura, con el objetivo de evitar los problemas referentes a la relación de perfecta linealidad existente entre las variables independientes, consiste en ignorar la variable cohorte (es decir, se asume que la tendencia lineal de esta variable toma un valor cero- efectos lineales y no lineales de este regresor serían considerados inexistentes). Esta estrategia posee un número importante de riesgos asociados; por lo que, se suele testar si un modelo que considera únicamente dos factores, permite obtener resultados similares a los derivados de la estimación de un modelo edad- periodo- cohorte. En caso de que esto ocurra, es recomendable utilizar el modelo con dos factores, en virtud del *principio de parsimonia*. Sin embargo, conviene tener en cuenta que, al obviar una de las variables independientes, la solución obtenida no forma parte de la línea de soluciones asociada al modelo edad- periodo- cohorte. En este aspecto, el modelo con dos componentes, absorbería cualquier componente lineal (y no lineal) de la variable independiente ignorada. En este caso, el grado de importancia de la tendencia lineal asociado a este tercer regresor, no podría ser analizado realizando, simplemente, una comparación entre el modelo con dos factores y el modelo que agrupa las tres variables independientes. En este aspecto, el modelo que considera un único factor adolece de los mismos inconvenientes; pudiendo originar, por tanto, problemas relacionados con la comprensión de los efectos asociados al único regresor considerado.

Por otra parte, resolver el modelo edad- periodo- cohorte mediante la imposición de una restricción, conlleva escoger una restricción adecuada. Para ello, se pueden ejecutar aquellas restricciones que resultan ser más coherentes desde un punto de vista teórico; aunque, se debe ser cauto en este aspecto, como señalan autores como O' Brien (2014) [104]; ya que, los resultados derivados de la imposición de cualquier restricción son igualmente válidos.

3.3.2.3. Geometría de los modelos edad- periodo- cohorte

Comprender, desde un punto de vista geométrico, los cimientos sobre los que se asientan los modelos edad- periodo- cohorte puede ser crucial a la hora de entender el problema de identificación de dichos modelos; así como, la estructura que subyace a estos. En este aspecto, se seguirá el procedimiento desarrollado por Strang (1998) [113], que considera a cada fila asociada a las ecuaciones normales, como un objeto geométrico contenido en un espacio de soluciones de m dimensiones. Por tanto, se estudiará, desde una perspectiva geométrica, la restricción utilizada a la hora de tratar de solventar el problema de identificación inherente a los modelos edad- periodo- cohorte.

De tal forma, desde un punto de vista geométrico, $m - 1$ hiperplanos, que representan a $m - 1$ de las ecuaciones normales, se cruzan entre sí para formar una línea en un espacio de dimensión m . En este aspecto, uno de los hiperplanos, que representa una de las ecuaciones normales, no se cruza en dicho punto; sino que, la línea formada por el resto de hiperplanos, reposa sobre este (originando infinitas soluciones, cuya obtención depende de la restricción adoptada).

Para ilustrar esta idea, podría representarse el modelo edad- periodo- cohorte en un espacio tridimensional. Para ello, se considerará que los regresores temporales del modelo están codificados como variables continuas a nivel de intervalo; estando centrados, a su vez, respecto a su media. Por tanto, en este caso, existirían tres hiperplanos, que representarían a las tres ecuaciones normales. En base al razonamiento expuesto anteriormente, dos de los planos asociados a dos de las ecuaciones normales se cruzarían formando una línea; aunque, el plano restante no se cruzaría, con dicha línea, en un único punto, sino que descansaría

sobre la misma; originando, por tanto, infinitas soluciones. Desde un punto de vista geométrico, esta línea podría ser determinada.

Para simplificar la exposición matemática de esta cuestión, se empleará la restricción de los modelos edad- periodo- cohorte más fácil de visualizar; siendo, esta, la que utiliza como variables continuas a los regresores edad, periodo y cohorte. Se utilizará, además, una codificación a nivel de intervalo. Por tanto, la ecuación asociada al modelo edad- periodo- cohorte será, en este caso:

$$y_{ij} = \lambda + b_{\alpha} \times \alpha_i + b_{\beta} \times \beta_j + b_{\gamma} \times \gamma_k + \epsilon_{ij} \quad (3.31)$$

siendo y_{ij} la ratio asociada al i -ésimo grupo de edad en el j -ésimo periodo. Además, λ es la constante del modelo y α_i es la edad correspondiente al i -ésimo grupo de edad. Por otro lado, β_j es el periodo asociado al j -ésimo periodo; mientras que, γ_k representa a la cohorte k . Finalmente, ϵ_{ij} es el residuo asociado al i -ésimo grupo de edad y al j -ésimo periodo. Por otro lado, conviene señalar que, la adición de otras variables independientes al modelo, no afectaría a la dependencia lineal existente entre los regresores edad, periodo y cohorte.

La geometría asociada a este caso revista de una mayor sencillez, en comparación con aplicaciones más complejas; aunque, son una base de comprensión para el resto de casos. En este aspecto, la relación existente entre los componentes lineales asociados a estos modelos puede reflejarse como se muestra a continuación (Rodgers, 1982)[114]:

$$\begin{aligned} b_{\alpha c}^0 &= b_{\alpha c1}^0 + s \\ b_{\beta c}^0 &= b_{\beta c1}^0 - s \\ b_{\gamma c}^0 &= b_{\gamma c1}^0 + s \end{aligned}$$

Esta notación ha sido extraída de O'Brien (2014; pp.70) [104]; considerando, en este caso, el caso centrado. De tal formas, los términos de la izquierda de dichas ecuaciones, se corresponden con las infinitas soluciones asociadas a los efectos lineales de los regresores; mientras que, los coeficientes de la derecha, hacen alusión a los efectos lineales de los coeficientes obtenidos a través de la imposición de la restricción c_1 . Igualmente, como se ha señalado en el resto del documento, s representa a un escalar. Se puede añadir que, a simple vista, existe una relación lineal entre los efectos descritos en estas tres ecuaciones.

Por otro lado, este conjunto de ecuaciones puede resumirse, de forma vectorial, mediante la siguiente expresión:

$$b_c^0 = b_{c1}^0 + sv \quad (3.32)$$

En este caso, b_c^0 refleja una de la infinitas soluciones asociadas al modelo; mientras que, b_{c1}^0 es la solución asociada a la restricción c_1 . Por otro lado, desde un punto de vista matricial, las expresiones introducidas anteriormente pueden expresarse de la siguiente forma:

$$\begin{bmatrix} b_{\alpha c}^0 \\ b_{\beta c}^0 \\ b_{\gamma c}^0 \end{bmatrix} = \begin{bmatrix} b_{\alpha c1}^0 \\ b_{\beta c1}^0 \\ b_{\gamma c1}^0 \end{bmatrix} + s \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad (3.33)$$

Por tanto, las infinitas soluciones correspondientes al modelo se encuentran en la línea $b_c^0 = b_{c1}^0 + sv$ (estas se encuentran contenidas en un espacio tridimensional). Para obtener esta línea se debe calcular, únicamente, una solución a las ecuaciones; añadiéndoles, posteriormente, el escalar sv .

La solución geométrica coincide con la solución obtenida a través de la implementación de métodos algebraicos. En este aspecto, debido a la imposibilidad de obtener la inversa de $X'X$, se debe calcular una matriz inversa generalizada que genere una solución de mínimos cuadrados asociada a las ecuaciones normales. Estos pasos quedan resumidos a través de la siguiente expresión:

$$(X'X)^- X'Xb = (X'X)^- X'y \quad (3.34)$$

La solución obtenida mediante la adopción de una matriz inversa generalizada $(X'X)_{c_1}^-$, asociada a la restricción c_1 , puede ser expresada de la siguiente manera:

$$b_{c_1} = (X'X)_{c_1}^- X'y \quad (3.35)$$

Desde un punto de vista geométrico, la matriz inversa generalizada permite restringir la dirección del plano (en base a la restricción adoptada); permitiendo obtener una solución, que será ortogonal respecto a la restricción. En este caso, se ha considerado un modelo edad- periodo- cohorte con codificación lineal; siendo, el caso de clasificación múltiple, el que reviste de una mayor complicación (debido a la existencia de un mayor número de dimensiones). En este aspecto, considerando m ecuaciones, existirían m dimensiones, en relación con el espacio de soluciones. Además, habrían $m \times (m - 1)$ hiperplanos dimensionales, asociados a las m ecuaciones normales. Por otro lado, la línea de soluciones estaría determinada, a su vez, mediante la intersección de un número de hiperplanos igual a $m - 1$, cuya expresión vendría dada en función de la ecuación 3.32 (a su vez, cada elemento de la ecuación, tendría un número de elementos igual a m). Finalmente, uno de los hiperplanos no poseería una intersección, en relación con la línea de soluciones; por lo que, se tendría que adoptar una restricción respecto a dicho hiperplano. De tal forma, esta restricción permitiría reorientar dicho hiperplano, con el objetivo de que este toque la línea de soluciones en un único punto (en lugar de en infinitos).

Por otro lado, el procedimiento geométrico posibilita la comprensión de las cuestiones del modelo asociadas a la solución con mínima norma. En este aspecto, la distancia existente entre cualquier solución correspondiente a la línea de soluciones del modelo y el punto de origen correspondiente al espacio multidimensional, toma el valor $\|b_{c_1}^0\|$. Por tanto, la solución de mínima norma se corresponde con el vector que presenta la menor distancia; siendo, esta, la solución que es perpendicular respecto al vector de ceros (solución *Moore- Penrose*). Por otra parte, las soluciones que se encuentran sobre la citada línea, estarán separadas por la distancia $|s| \times \|v\|$.

Por otro lado, al usar regresiones restringidas a la hora de resolver los modelos edad- periodo- cohorte, es conveniente elegir una restricción que se adecue al patrón de generación de los datos. Sin embargo, existen procedimientos mecánicos que aplican restricciones idénticas; independientes, pues, del patrón de generación de los datos. Este procedimiento generará soluciones insesgadas, únicamente, en los casos en los que la restricción adoptada esté asociada con el procedimiento a través del cual se generan los valores asociados a la variable respuesta. En este aspecto, la solución derivada de la imposición de una restricción mecánica podría ser más eficiente, en comparación con las soluciones asociadas a otras restricciones, desde el punto de vista de la minimización de los errores estándar o de la maximización de la entropía. De tal forma, procedimientos como los adoptados a la hora de implementar el estimador intrínseco, emplean restricciones que no se encuentran asociadas al procedimiento de generación de los datos (en este caso, para encontrar una solución del modelo, se aplica una restricción de tipo *Moore- Penrose*, con la matriz inversa generalizada correspondiente). De hecho, autores como Yang et al. (2008) [118] hacen referencia a las propiedades que subyacen al estimador intrínseco (mínima varianza, menor longitud del vector de soluciones en relación con la distancia existente entre el origen y la línea de soluciones, perpendicularidad de dichas soluciones respecto al vector de ceros, ...). Por tanto, en este caso, el investigador no se vería obligado a justificar la adopción de esta restricción, debido a las importantes ventajas que esta posee desde un punto de vista estadístico. Sin embargo, es conveniente analizar si, la restricción asociada a la matriz inversa generalizada de *Moore- Penrose* presenta, desde el punto de vista de la teoría asociada al estudio, cierta coherencia.

3.3.3. El papel de las funciones estimables

Como se explicó anteriormente, al aplicar el procedimiento de imposición de restricciones, existe un número infinito de posibles soluciones, igualmente válidas, relacionadas con los coeficientes de las variables edad, periodo y cohorte. En este aspecto, no se puede escoger la mejor de estas estimaciones a partir del estudio de la verosimilitud correspondiente a los coeficientes estimados; ya que, este valor es igual en todos los casos. Sin embargo, existen métodos que permiten obtener estimaciones, no sesgadas, relacionadas con las combinaciones lineales asociadas a los datos. En este contexto, destacan las funciones estimables, cuyo valor es independiente de la restricción adoptada. De hecho, las funciones estimables se caracterizan por ser combinaciones lineales de las soluciones que se encuentran sobre la línea $b_c^0 = b_{c_1}^0 + sv$. En este procedimiento,

el vector nulo posee un papel fundamental, debido a sus características intrínsecas.

En primer lugar, en el contexto de los modelos edad- periodo- cohorte, una función estimable consiste en una combinación lineal, de los parámetros, que permite generar los valores correspondientes a la variable respuesta. En este aspecto, si se combinan las posibles soluciones del modelo a través de las restricciones correspondientes, se podría calcular una estimación insesgada de dichos parámetros (O' Brien, 2014) [104]. De tal forma, las funciones estimables son invariantes ante cualquier solución obtenida a partir de la resolución de las ecuaciones normales; por lo que, el cálculo de estas, reviste de un especial interés. En este aspecto, Searle (1971; pp. 160- 187) [116] estableció una condición necesaria y suficiente para que, una combinación lineal de las soluciones que se encuentre sobre la línea de soluciones ($q'b_c^0$), sea considerada una función estimable; definiendo, previamente, el valor de H ($H = GX'X$), donde G es, a su vez, la matriz inversa generalizada. De tal forma, para que la combinación lineal $q'b_c^0$ sea considerada una función estimable, se debe cumplir lo siguiente:

$$q'H = q' \quad (3.36)$$

En este caso, q' es un vector que, al ser multiplicado por el vector de soluciones, genera una función estimable (dicha función es, a su vez, una combinación lineal del vector de soluciones).

Tras esta breve explicación, se procederá a introducir uno de los múltiples métodos que permite calcular estas funciones; denominándose, este último, *aproximación l'sv*. Dicho procedimiento se encuentra recogido en O' Brien (2012) [117] y parte del hecho de que, cualquier solución de los modelos edad- periodo- cohorte, se encuentra contenida sobre una línea de soluciones ($b_c^0 = b_{c1}^0 + sv$) situada en un espacio de m dimensiones. En este aspecto, las funciones estimables son combinaciones lineales de las soluciones que se encuentran situadas sobre dicha línea; aunque, para que una combinación lineal sea considerada una función estimable, esta debe ser invariante respecto a las diferentes restricciones que pueden ser ejecutadas a la hora de obtener las soluciones.

El procedimiento *l'sv* parte del hecho de que, todas las soluciones que se encuentran sobre la línea $b_c^0 = b_{c1}^0 + sv$, difieren entre sí en función de sv . De tal forma, si se toma una combinación lineal, l , sobre esta línea, se tiene que $l'b_c^0 = l'b_{c1}^0 + l'sv$. El objetivo de este método consiste en calcular las combinaciones lineales $l'b_c^0$ y $l'b_{c1}^0$ que no dependen del valor de s . Esto ocurre, únicamente, si $l'sv = 0 \forall s$. De tal forma, la combinación lineal debe ser la misma para todas las soluciones situadas sobre la línea de soluciones.

Partiendo de este hecho, se podría considerar al *estimador intrínseco (IE)* como una función estimable, de acuerdo a las sugerencias de autores como Yang et al. (2008) [118]. Sin embargo, autores como O' Brien (2014) [104] (pp. 98) sostienen que es confuso considerar a este estimador como una función estimable; ya que, los elementos de su vector de soluciones no son funciones estimables, desde un punto de vista individual.

3.3.3.1. Aproximación *l'sv* y obtención de funciones estimables

Coefficientes de efecto

En el contexto de los modelos de clasificación edad- periodo- cohorte múltiples, algunos coeficientes de efectos de índole individual pueden ser calculados. De tal forma, partiendo de que las infinitas soluciones del modelo se sitúan sobre la línea de soluciones (diferenciándose en función de sv), se debe examinar el vector de ceros asociado a cada uno de los efectos. En este aspecto, si el vector nulo es igual a cero, independientemente de la solución obtenida y de la restricción adoptada, el parámetro obtenido no presentará variación alguna a lo largo de la línea de soluciones.

Desde un punto de vista matemático, siendo v_i el elemento i -ésimo del vector de ceros y b_i el elemento i -ésimo del vector de soluciones, si se tiene que $v_i = 0$, se cumpliría lo siguiente (independientemente del valor de s): $b_i + sv_i = b_i$. Por tanto, en este caso se estaría ante una función estimable; aunque, dichas funciones no poseen una notable relevancia, desde un punto de vista interpretativo.

Segundas diferencias

El cálculo de las segundas diferencias reviste de una importancia fundamental en los modelos edad-periodo-cohorte (como se explicó en el epígrafe 3.3.1); ya que, permite analizar el aumento, o la disminución, asociada a la ratio de variaciones correspondiente a los coeficientes edad, periodo y cohorte. Además, se debe tener en cuenta que, el valor de las segundas diferencias, no se ve influenciado por la elección de la restricción utilizada para resolver el modelo. En este aspecto, se considerará que $v_{i\alpha}$ es el elemento del vector de ceros que se encuentra relacionado con el efecto de la variable edad α_i (siendo $i = 1$ hasta I). Si se tienen en cuenta los tres primeros efectos asociados al regresor edad, se pueden considerar las segundas diferencias correspondientes a esta variable de la siguiente forma (desde el punto de vista de la aproximación $l'sv$):

$$(\alpha_3 - \alpha_2) - (\alpha_2 - \alpha_1) = \alpha_3 + \alpha_1 - 2\alpha_2$$

Si se considera una de las soluciones pertenecientes a la línea de soluciones (mediante la adición del escalar s a la expresión introducida anteriormente), se tiene la segunda derivada siguiente:

$$[(\alpha_3 + sv_{3\alpha}) - (\alpha_2 + sv_{2\alpha})] - [(\alpha_2 + sv_{2\alpha}) + (\alpha_1 + sv_{1\alpha})] =$$

Operando se obtiene, finalmente, la siguiente expresión:

$$= \alpha_1 - 2\alpha_2 + \alpha_3 + (sv_{1\alpha} - 2sv_{2\alpha} + sv_{3\alpha})$$

donde $(sv_{1\alpha} - 2sv_{2\alpha} + sv_{3\alpha}) = 0$; ya que, el elemento $v_{1\alpha}$ es una variable lineal codificada en un intervalo fijo. Por tanto, para cualquier nueva solución, la segunda diferencia puede calcularse mediante la expresión $\alpha_1 - 2\alpha_2 + \alpha_3$, que es idéntica a las segundas diferencias pertenecientes a la solución original del modelo.

Por otro lado, dichos resultados pueden ser generalizados mediante un cambio de notación en los subíndices. En este aspecto, α pasaría a ser α_i, α_{i+1} y α_{i+2} ; mientras que, v sería igual a $v_{i\alpha}, v_{(i+1)\alpha}$ y $v_{(i+2)\alpha}$, según la notación utilizada en O' Brien (2014) [104]. Por otra parte, las segundas diferencias para los periodos y las cohortes, podrían ser calculadas siguiendo un procedimiento análogo al que se acaba de introducir para la variable edad.

Relación existente entre las pendientes

El término sv , unido a las pendientes de los regresores del modelo edad-periodo-cohorte, puede resultar de gran relevancia a la hora de llevar a cabo el proceso de interpretación del modelo. De tal forma, con el objetivo de calcular la variación de los coeficientes asociados a las pendientes, al cambiar de una solución a otra, es conveniente hacer uso de la combinación sv .

En este aspecto, la variación de la pendiente, asociada a la variable edad, que se produce al incorporar el término $sv_{i\alpha}$ a los coeficientes α_i , es igual a $s(v_{2\alpha} - v_{1\alpha})$. Igualmente, si se incorpora un valor $sv_{j\beta}$ a cada uno de los coeficientes de la variable periodo, la pendiente de este regresor se ve modificada en una cantidad determinada por $s(v_{2\beta} - v_{1\beta})$; mientras que, si se añade una cantidad igual a $sv_{k\gamma}$ a cada coeficiente de la variable cohorte, la pendiente se vería modificada en $s(v_{2\gamma} - v_{1\gamma})$.

Por otro lado, se puede considerar que primera solución obtenida posee una tendencia t_α , en relación con los coeficientes de la variable edad, t_β para los coeficientes del regresor periodo y t_γ para los coeficientes de la variable cohorte. Así, y considerando que cualquier nueva solución se encuentra situada sobre la línea de soluciones (y difieren entre sí, pues, en función de sv), se tiene que la pendiente de la nueva solución, en relación con las variables edad, periodo y cohorte; posee las siguientes expresiones, respectivamente:

$$t_\alpha + s(v_{2\alpha} - v_{1\alpha})$$

$$t_\beta + s(v_{2\beta} - v_{1\beta})$$

$$t_\gamma + s(v_{2\gamma} - v_{1\gamma})$$

Además, debido a las característica del vector nulo (en relación con su unicidad), se cumple lo siguiente:

$$|v_{2\alpha} - v_{1\alpha}| = |v_{2\beta} - v_{1\beta}| = |v_{2\gamma} - v_{1\gamma}|$$

es decir, las distancias existentes entre los elementos de los vectores nulos correspondiente a los distintos regresores del modelo poseen el mismo valor, en términos absolutos.

Estas relaciones están presentes en las funciones estimables de las pendientes del modelo. En este aspecto, la suma de las tendencias de las variables edad y periodo puede ser estimada; presentado, esta última, independencia en relación con la restricción adoptada. De tal forma, se debe partir del hecho de que $(v_{2\alpha} - v_{1\alpha})$ y $(v_{2\beta} - v_{1\beta})$ poseen el mismo valor (en términos absolutos); aunque, presentan distinto signo. Por tanto, como $s(v_{2\alpha} - v_{1\alpha}) + s(v_{2\beta} - v_{1\beta}) = 0$ se tiene que:

$$t_{\alpha} + t_{\beta} = (t_{\alpha} + s(v_{2\alpha} - v_{1\alpha})) + (t_{\beta} + s(v_{2\beta} - v_{1\beta}))$$

En este aspecto, $t_{\beta} + t_{\gamma}$ también pueden estimarse, siguiendo un procedimiento idéntico al que se acaba de introducir. Por tanto, como $s(v_{2\beta} - v_{1\beta}) + s(v_{2\gamma} - v_{1\gamma}) = 0$ se tiene que:

$$t_{\beta} + t_{\gamma} = (t_{\beta} + s(v_{2\beta} - v_{1\beta})) + t_{\gamma} + (s(v_{2\gamma} - v_{1\gamma}))$$

Igualmente, para la diferencia de tendencias asociada a las variables edad y cohorte, se tiene lo siguiente; ya que, $s(v_{2\alpha} - v_{1\alpha}) - s(v_{2\gamma} - v_{1\gamma}) = 0$:

$$t_{\alpha} - t_{\gamma} = (t_{\alpha} + s(v_{2\alpha} - v_{1\alpha})) - (t_{\gamma} + s(v_{2\gamma} - v_{1\gamma}))$$

Partiendo de estas cuestiones, se pueden plasmar las relaciones existentes entre las pendientes de los regresores. De tal forma, considerando que:

$$t_{\alpha} + t_{\beta} = k_1$$

$$t_{\beta} + t_{\gamma} = k_2$$

$$t_{\alpha} - t_{\gamma} = k_3$$

se tiene que $(t_{\alpha} + t_{\beta}) - (t_{\alpha} + t_{\gamma}) = k_1 - k_2 = k_3$. Por tanto, partiendo de una serie de funciones estimables, puede obtenerse otra función estimable.

Finalmente, autores como Holford (1985) [119] u O' Brien (2014) [104] sostienen que, la expresión que permite reflejar la estimación de la relación existente entre las pendientes, puede ser representada de la siguiente manera:

$$d_1 t_{\alpha} + d_2 t_{\beta} + (d_2 - d_1) t_{\gamma} \tag{3.37}$$

De tal forma, mediante el uso del método *l'sv*, esta expresión puede ser reescrita, con el objetivo de que se mantenga constante con independencia de la restricción adoptada. En este aspecto, una nueva solución asociada al modelo edad- periodo- cohorte puede ser escrita de la siguiente manera: $d_1(t_{\alpha} + s(v_{2\alpha} - v_{1\alpha})) + d_2(t_{\beta} + s(v_{2\beta} - v_{1\beta})) + (d_2 - d_1)(t_{\gamma} + s(v_{2\gamma} - v_{1\gamma}))$. Operando, se obtiene, finalmente, lo siguiente:

$$d_1 t_{\alpha} + d_2 t_{\beta} + (d_2 - d_1) t_{\gamma}$$

Esta función puede ser estimada, presentando independencia en relación con la restricción adoptada.

Variación en la pendiente considerando las modificaciones de los factores propios

Las variaciones en las pendientes dentro de los grupos de edad, periodo y cohorte pueden ser estimadas. En este aspecto, suponiendo que se han calculado dos tendencias lineales ($t_{\gamma 1}$ y $t_{\gamma 2}$), en relación con dos conjuntos de coeficientes de la variable cohorte; se tiene que, el cambio en la pendiente, puede ser calculado mediante la siguiente operación: $t_{\gamma 2} - t_{\gamma 1}$. Si se obtiene una nueva solución, la tendencia para el primer conjunto de coeficientes viene dada por la expresión $t_{\gamma 1} + s(v_{2\gamma} - v_{1\gamma})$; mientras que, la tendencia asociada al segundo conjunto de coeficientes, sería $t_{\gamma 2} + s(v_{2\gamma} - v_{1\gamma})$. Por tanto, esta segunda solución poseería una variación de sus pendientes igual a:

$$t_{\gamma 2} - t_{\gamma 1} = (t_{\gamma 2} + s(v_{2\gamma} - v_{1\gamma})) - (t_{\gamma 1} + s(v_{2\gamma} - v_{1\gamma})) = t_{\gamma 2} - t_{\gamma 1}$$

De tal forma, el valor de la variación en las pendientes es independiente de la solución obtenida (y, por tanto, de la restricción adoptada). Esta explicación puede extenderse a las tendencias asociadas a las variables edad y periodo.

Modificaciones en la parte lineal del modelo

En primer lugar, se debe tener en cuenta que, los componentes lineales correspondientes a los modelos edad- periodo- cohorte, no pueden ser estimados; aunque, se podrían calcular los elementos no lineales asociados a las variables independientes, como se explicó anteriormente. En este aspecto, como se señala en O' Brien (2014; pp. 103) [104], las variaciones de los coeficientes de las variables edad, periodo y cohorte, en relación con la tendencia lineal de dichos coeficientes, presentan la propiedad de invarianza. Para demostrar esta cuestión, se partirá del vector α_i , que representa los coeficientes asociados a la variable edad, en relación con la solución original. Además, $v_{i\alpha}$ hace referencia a los elementos contenidos, en el vector de ceros, correspondientes a dichos coeficientes; mientras que, t_α hace alusión a la tendencia obtenida en la solución original. Considerando el modelo centrado, se tiene que los valores centrados toman el valor i_α ; por lo que, los valores de los coeficientes asociados a la variable edad, en relación con los efectos lineales, son $t_\alpha i_\alpha$. Por tanto, el término $\alpha_i - t_\alpha i_\alpha$ hará referencia a las desviaciones existentes, en relación con la linealidad, asociadas a los efectos de la variable edad.

De tal forma, se supondrá que se obtiene otra solución para los efectos de la variable edad; poseyendo, esta última, la forma $\alpha_i + sv_{i\alpha}$. Por tanto, la tendencia asociado a dicho coeficiente tomará un valor de $t_\alpha + s(v_{2\alpha} - v_{1\alpha})$; mientras que, el valor predicho de este, será $i_\alpha \times (t_\alpha + s(v_{2\alpha} - v_{1\alpha}))$ (considerando, únicamente, los efectos lineales). Por ende, las desviaciones asociadas a las distintas soluciones están representadas por la expresión $[\alpha_i + sv_{i\alpha}] - [(t_\alpha + s(v_{2\alpha} - v_{1\alpha})) \times i_\alpha]$. En este aspecto, como $sv_{i\alpha} = s(v_{2\alpha} - v_{1\alpha}) \times i_\alpha$, la desviación original será igual a la nueva desviación (derivada, esta última, de la adopción de una nueva restricción a la hora de resolver el modelo):

$$\alpha_i - t_\alpha i_\alpha = [\alpha_i + sv_{i\alpha}] - [(t_\alpha + s(v_{2\alpha} - v_{1\alpha})) \times i_\alpha]$$

Por tanto, se pueden estimar las divergencias, respecto a los efectos en la variable edad, asociadas a la tendencia lineal (estas son, además, funciones estimables). Este procedimiento puede extenderse, a su vez, a las variables periodo y cohorte.

Valores predichos de la variable de interés, y

Como se señaló anteriormente, los valores predichos de la variable de interés (y) toman valores idénticos, independientemente de la restricción adoptada. Esta circunstancia está causada por una de las propiedades del vector v ; que puede ser representada, a su vez, a través de la siguiente expresión:

$$v_0 + v_{i\alpha} + v_{j\beta} + v_{k\gamma} = 0$$

reflejando v_0 la constante, $v_{i\alpha}$ el i -ésimo grupo asociado a la variable edad, $v_{j\beta}$ el j -ésimo periodo y $v_{k\gamma}$ la k -ésima cohorte. De tal forma, para cualquier celda asociada a una tabla en la que se reflejen dos de las variables independientes del modelo; se tiene que, las sumas de los elementos de los vectores nulos asociados a los efectos de las variables edad, periodo y cohorte (sumando la constante), toman un valor igual a 0.

Por tanto, considerando un tabla que refleje las variables edad y periodo; se tiene que, el valor predicho para la ij -ésima celda de dicha tabla puede calcularse a través de la siguiente expresión:

$$\hat{y}_{ij} = \text{constante} + \alpha_i + \beta_j + \gamma_k \quad (3.38)$$

Por tanto, una nueva solución contenida sobre la línea de soluciones, puede originar los siguientes valores predichos:

$$\hat{y}_{ij} = (\text{constante} + sv_0) + (\alpha_i + sv_{i\alpha}) + (\beta_j + sv_{j\beta}) + (\gamma_k + sv_{k\gamma}) \quad (3.39)$$

Agrupando términos, esta expresión puede reescribirse de la siguiente manera: $\hat{y}_{ij} = \text{constante} + \alpha_i + \beta_j + \gamma_k + s(v_0 + v_{i\alpha} + v_{j\beta} + v_{k\gamma})$. Por tanto, los valores predichos son iguales en ambas soluciones; ya que,

$$v_0 + v_{i\alpha} + v_{j\beta} + v_{k\gamma} = 0.$$

En términos generales, existe un número importante de funciones estimables que pueden ser obtenidas a través de la aproximación *l'sv*; siendo, este procedimiento, especialmente flexible.

3.3.4. Métodos de identificación bayesianos

Por otro lado, algunos autores han abandonado la metodología frecuentista con el objetivo de estudiar los modelos edad- periodo- cohorte, abogando por el uso de métodos de identificación bayesianos. Sin embargo, esta última metodología presenta una serie de problemas similares a los que se obtenían al tratar de solucionar el problema de identificación a través de la imposición de restricciones.

Para explicar los fundamentos inherentes a los métodos de identificación bayesianos, se seguirán los postulados contenidos en Fannon y Nielsen (2018) [91]. En este aspecto, la función de verosimilitud está condicionada por el vector de los efectos temporales, θ (ver ecuación 3.7); por lo que, posee la siguiente expresión: $p(Y|\theta)$, donde Y hace alusión a la matriz de datos. De tal forma, la distribución a priori es $p(\theta)$; por lo que, el parámetro θ puede descomponerse de la siguiente manera:

$$\theta = (\xi, \varrho)$$

donde ξ se corresponde con el parámetro canónico (que posea dimensión p); mientras que, ϱ hace referencia a la parte de θ que no puede ser identificada. Por tanto, se cumplirá lo siguiente:

$$p(Y|\theta) = p(Y|\xi)$$

puediendo descomponerse, la distribución a priori, de la siguiente manera:

$$p(\xi, \varrho) = p(\xi) \times p(\varrho|\xi)$$

siendo $p(\xi)$ la distribución a priori para el parámetro identificable y $p(\varrho|\xi)$ la distribución a priori para el parámetro no identificable condicionada al parámetro identificable, ξ . Por otro lado, la distribución a posteriori presenta la siguiente expresión:

$$p(\theta|Y) = p(\xi|Y) \times p(\varrho|\xi, Y)$$

por lo que, en base a Poirier (1998) [110]:

$$p(\xi|Y) = p(Y|\xi)p(\xi)/p(Y) \text{ y } p(\varrho|\xi, Y) = p(\varrho|\xi) \quad (3.40)$$

Por tanto, de acuerdo con esta última expresión se observa que, la función de verosimilitud, se encarga de actualizar el parámetro canónico; aunque, no posee la capacidad de actualizar la información, a priori, relacionada con el parámetro ϱ (que era, a su vez, no identificable). Por tanto, este problema es similar al que se obtenía a la hora de resolver el problema de identificación del modelo mediante la imposición de restricciones en la parte lineal del mismo.

Por otra parte, en los modelos edad- periodo- cohorte bayesianos, se suele escoger una distribución a priori que se encargue de generar dobles diferencias independientes e idénticamente normales. En este aspecto, como se señala en Fannon y Nielsen (2018) [91], autores como Berzuini y Clayton (1994) [111] abogan por emplear una distribución a priori uniforme, en la que se incluya el parámetro ϱ , con el objetivo de representar las partes lineales asociadas a los efectos temporales. Sin embargo, autores como Smith y Wakefield (2016) [112], consideran que es más conveniente utilizar una distribución a priori que no incorpore el parámetro ϱ a la hora de representar el plano lineal del modelo. Por tanto, dichos autores sostienen que, no se debe suponer una distribución, a priori, para el parámetro que no puede identificarse.

3.3.5. Conclusiones respecto al problema de identificación

Para cerrar este epígrafe, es conveniente resumir algunas de las cuestiones principales asociadas al problema de identificación de los modelos edad- periodo- cohorte. En primer lugar, dicho problema se originaba debido a la multicolinealidad existente en los regresores temporales del modelo (edad, periodo y cohorte); provocando que, las partes lineales de este, no pudieran ser identificadas. Sin embargo, se pueden calcular las partes no lineales referentes a los efectos temporales; además de, el plano lineal sobre el que se asienta el predictor (μ_{ij} , considerando un *age- period array*). Desde un punto de vista práctico, es fundamental considerar los efectos lineales y no lineales de forma separada.

Por otro lado, en cuanto a los métodos que tratan de solventar el problema de identificación, se han explicado los inconvenientes asociados a utilizar la imposición de restricciones referentes al plano lineal. En este aspecto, esta metodología genera una serie de problemas relacionados con la interpretación, la formulación de hipótesis o la imposibilidad de reflejar los componentes no lineales del predictor. De igual forma, los métodos de identificación bayesianos generan consecuencias similares a las obtenidas mediante la imposición de restricciones. Finalmente, la parametrización canónica permite considerar los efectos lineales (mediante un plano lineal) y no lineales (a través de dobles diferencias) del modelo; por lo que, posee una serie de ventajas relacionadas con la estimación y la formulación de hipótesis.

3.4. Estimación del modelo

En un número importante de casos, los modelos edad- periodo- cohorte presentan una modelización que se basa en los modelos lineales generalizados (GLM). Este es el caso, por ejemplo, del modelo Poisson dosis-respuesta (*Poisson dose-response model*), que será utilizado en el proceso de análisis de este estudio. En este aspecto, los datos subyacentes a este tipo de modelos se basan en:

- Frecuencias. Estas se reflejan en base al elemento Y_{ik} .
- Dosis o exposiciones (*doses or exposures*). Estas serán denotadas en base al elemento Z_{ij} .

Por tanto, las variables edad y periodo (se está suponiendo, pues, la utilización de un *age- period array*) se desplazan en un trapecioide generalizado; mientras que, las respuestas (Y_{ij}) se distribuyen, de forma independiente, en base a una distribución de Poisson. Además, las respuestas se encuentran condicionadas a las dosis o exposiciones. Por estas razones, el valor esperado posee la siguiente estructura:

$$E(Y_{ij}|Z_{ij}) = \exp(\mu_{ij})Z_{ij} = \exp(\mu_{ij} + \log(Z_{ij})) \quad (3.41)$$

Por tanto, el logaritmo de verosimilitud poseerá la siguiente expresión:

$$l(\xi; Y|Z) = \sum_{i,j \in I} Y_{ij} (\mu_{ij} + \log(Z_{ij})) - \sum_{i,j} Z_{ij} \exp(\mu_{ij}) \quad (3.42)$$

Si se sustituye el predictor, μ_{ij} , por los efectos temporales, se obtendrá una función de verosimilitud que no puede ser identificada; ya que, la matriz de diseño, tendrá rango deficiente en uno.

Por otro lado, la función de verosimilitud es utilizada para estimar el modelo; sirviendo, a su vez, para obtener el *criterio de información de Akaike* (AIC). En este aspecto, el *criterio de Akaike* es especialmente utilizado en el análisis de desviaciones; ya que, permite escoger el modelo (de entre todos los submodelos) que presenta una mejor adecuación en relación con la base de datos que se está analizando. En este aspecto, el *criterio de información de Akaike* únicamente puede ser utilizado como elemento de comparación de modelos. De tal forma, este estadístico se obtiene a partir de la siguiente expresión:

$$\text{AIC} = -2\log L + 2 \times p \quad (3.43)$$

siendo L el valor de la función de máxima verosimilitud; mientras que, p se corresponde con el número de parámetros existente en el modelo.

3.5. Descomposición de la varianza y extensiones

Se puede realizar una descomposición de la varianza mediante la metodología de análisis de la varianza; siendo, el resultado de este procedimiento, una función estimable. Este proceso está estrechamente relacionado con los *modelos edad- periodo- cohorte mixtos (APCMM models)*; ya que, estos modelos permiten calcular la varianza referente a los efectos aleatorios (sin embargo, no posibilitan la obtención de funciones estimables).

Por otro lado, como se ha señalado a lo largo del documento, si se incorporasen, a la hora de estimar el modelo, dos de las variables independientes (en lugar de los tres regresores típicos), estas dos variables absorberían los efectos lineales de la variable temporal excluida. De tal forma, lo único que quedaría en este modelo, respecto a dicho regresor, serían las desviaciones de los coeficientes de este en relación con la linealidad. En este aspecto, las varianzas de estas desviaciones pueden ser estimadas; utilizando, para ello, una partición de la varianza. A continuación, se estudiarán una serie de métodos que permiten realizar esta técnica; destacando, en este sentido, los siguientes:

- *Análisis edad- periodo- cohorte de la varianza.* Se aplicará, en este aspecto, la técnica *ANOVA*; considerando, además, el caso de de la aplicación de esta técnica a los submodelos.
- Aplicación de extensiones de los modelos edad- periodo- cohorte. Destacan, en este caso, los *modelos edad- periodo- cohorte mixtos (APCMM models)* y la *aproximación HAPC* (Yang y Land, 2006) [120].

En este aspecto, las técnicas *ANOVA* y los *modelos APCMM*, permiten obtener una condición suficiente en relación con la aportación de cada variable independiente a la bondad del ajuste; por lo que, posibilitan la determinación de la significatividad de los regresores, desde un punto de vista estadístico.

3.5.1. Análisis edad- periodo- cohorte de la varianza (aproximación ANOVA)

El método de análisis de la varianza (ANOVA), es especialmente utilizado en los modelos edad- periodo- cohorte que son estimados a través del procedimiento de mínimos cuadrados ordinarios (OLS). En este aspecto, el coeficiente de correlación múltiple al cuadrado, R^2 , para el modelo completo será representado a través de la siglas R_{apc}^2 . De tal forma, el coeficiente de correlación múltiple al cuadrado muestra la proporción de la varianza total de la variable dependiente (y) que es explicada por el modelo estimado. Por otro lado, si en un modelo se consideran, únicamente, dos regresores (edad y periodo, por ejemplo), se representará el coeficiente de correlación múltiple al cuadrado a través de la siguiente expresión R_{ap}^2 . Sin embargo, a pesar de que la variable cohorte no está incluida en el modelo, sus efectos lineales estarán presentes en el mismo, debido a la dependencia lineal existente entre las variables edad, periodo y cohorte. Por tanto, la proporción de la varianza de la variable respuesta asociada a la cohorte, podrá calcularse a partir de la siguiente expresión: $R_{diff}^2 = R_{apc}^2 - R_{ap}^2$. De hecho, esta proporción de la varianza haría referencia a los efectos no lineales correspondientes a la variable cohorte. Este proceso podría extrapolarse, de forma equivalente, a las variables edad y periodo. Por otro lado, dicha expresión también puede usarse en modelos lineales generalizados (Poisson, por ejemplo); aunque, para comparar la bondad del ajuste entre distintos modelos, debe utilizarse el estadístico de razón de verosimilitud chi- cuadrado, que posee la siguiente expresión:

$$- 2 * \left(\frac{\text{Logaritmo de la función de verosimilitud para el modelo reducido}}{\text{Logaritmo de la función de verosimilitud para el modelo completo}} \right) \quad (3.44)$$

El resultado arrojado por esta última expresión sigue, aproximadamente, una distribución chi- cuadrado, con una cifra de grados de libertad igual a la diferencia existente entre el número de variables del modelo completo y la cifra de regresores del modelo restringido. De tal forma, este procedimiento es análogo al test R^2 (aunque, en este caso, no se podría hablar de varianza explicada por el modelo); siendo utilizado, este último, a la hora de comparar modelos que son estimados mediante mínimos cuadrados ordinarios. Por otro lado, también se pueden utilizar los *criterios de información* a la hora de comparar las diferencias existentes (en términos de la bondad del ajuste) entre los distintos modelos estimados; destacando, en este aspecto, el *criterio de Información Bayesiano (BIC)* y el *criterio de Información de Akaike (AIC)*.

3.5.2. Submodelos

La literatura de los modelos edad- periodo- cohorte presenta una serie de submodelos que revisten de un especial interés. En este aspecto, dichos submodelos pueden ser vistos como restricciones lineales respecto al parámetro canónico. Por tanto, como dichas restricciones poseen un carácter lineal, los submodelos surgidos presentan, al igual que el modelo general, las propiedades de las familias exponenciales regulares; facilitando, de esta manera, el proceso de estimación, el cálculo de los grados de libertad, la resolución de la prueba de razón de verosimilitud o la determinación de las desviaciones (*deviances*).

Por otro lado, el modelo edad- periodo- cohorte y el resto de submodelos restringidos se encuentran anidados; por lo que, se puede realizar una comparación entre estos a partir de un análisis de desviaciones (*deviances*). Para ello, se puede utilizar el *estadístico chi- cuadrado*.

Las actividades comparativas entre estos modelos también pueden realizarse empleando el *criterio de información de Akaike* (AIC), cuya expresión fue introducida en la ecuación 3.43. En este caso, se debería escoger el modelo (o submodelo) con el *criterio de información de Akaike* más reducido.

A continuación, se introducirán los principales submodelos asociados a los modelos edad- periodo- cohorte; haciendo alusión, a su vez, a las características más destacadas de estos.

3.5.2.1. Modelos edad- cohorte

Este tipo de submodelos es especialmente utilizado en el campo de la economía, como señalan autores como Fannon y Nielsen (2018) [91]; surgiendo, este último, al imponer una serie de restricciones sobre el modelo edad- periodo- cohorte. En este aspecto, partiendo del predictor asociado al modelo general:

$$\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta$$

el modelo edad- cohorte (*AC model*), surgiría al imponer las siguientes restricciones sobre el predictor:

$$\beta_{L+1} = \dots = \beta_{L+P} = 0$$

Por tanto, el predictor introducido anteriormente tomaría la siguiente forma:

$$\mu_{ik} = \alpha_i + \gamma_k + \delta \tag{3.45}$$

Desde un punto de vista paramétrico; podría señalarse que, la restricción impuesta en la expresión del predictor, presenta P dimensiones. Sin embargo, esta afirmación sería errónea; ya que, se ha impuesto una restricción que fija los efectos no lineales del regresor periodo a cero; por lo que, desde el punto de vista de la parametrización canónica, la restricción poseería $P - 2$ dimensiones, teniendo asociada, a su vez, la siguiente expresión:

$$\Delta^2 \beta_{L+3} = \dots = \Delta^2 \beta_{L+P} = 0$$

Por otro lado, autores como Nielsen y Nielsen (2014) [101] señalan que esta última restricción no consigue restringir, completamente, el plano lineal del modelo. Por tanto, los efectos lineales asociados a la variable periodo seguirán estando presentes en la expresión 3.45. En este aspecto, dichos efectos lineales no son eliminados del modelo (a pesar de la imposición de la restricción); ya que, son absorbidos por las pendientes de los regresores edad y cohorte, como se señaló en el epígrafe anterior.

Finalmente, en este contexto, se puede aplicar la aproximación ANOVA, con el objetivo de calcular los efectos no lineales asociados a la variable cohorte.

3.5.2.2. Submodelos de tipo lineal

En el cuadro 3.1 se introducen una serie de submodelos de especial relevancia en la metodología de los modelos edad- periodo- cohorte.

Cuadro 3.1: Grados de libertad de los distintos submodelos asociados a los modelos edad- periodo- cohorte. Fuente: Nielsen (2015) [129].

Tipo de modelo	Plano lineal	Dobles diferencias			Total
		$\Delta^2\alpha_i$	$\Delta^2\beta_j$	$\Delta^2\gamma_k$	
<i>APC (unrestricted)</i>	3	A-2	P-2	C-2	A+P+C-3
<i>AP</i>	3	A-2	P-2		A+P-1
<i>AC</i>	3	A-2		C-2	A+C-1
<i>PC</i>	3		P-2	C-2	P+C-1
<i>Age- Drift</i>	3	A-2			A+1
<i>Period- Drift</i>	3		P-2		P+1
<i>Cohort- Drift</i>	3			C-2	C+1
<i>A</i>	2	A-2			A
<i>P</i>	2		P-2		P
<i>C</i>	2			C-2	C
<i>Linear plane</i>	3				3

En relación con este cuadro, aparece, en la primera fila, el modelo edad- periodo- cohorte general (*APC unrestricted*); es decir, sin restricciones. Por otro lado, los modelos *AP*, *AC* y *PC* poseen una restricción, sobre las dobles diferencias, referente a las variables cohorte, periodo y edad, respectivamente. Por tanto, los efectos no lineales asociados a la variable no considerada, han sido eliminados a través de una restricción sobre las dobles diferencias. De tal forma, las conclusiones señaladas en el epígrafe anterior (modelos edad- cohorte) pueden ser aplicadas, con sus respectivas variaciones, a dichos submodelos.

Por otra parte, el submodelo *Period- Drift*, por ejemplo, posee las dobles diferencias de las variables edad y cohorte fijadas a cero. Sin embargo, el plano lineal del modelo no se encontraría restringido, como se señaló en anteriores apartados. Por tanto, el problema de identificación referente a los modelos edad- periodo- cohorte seguiría estando presente, debido a la multicolinealidad asociada a los regresores del modelo. En este aspecto, los modelos *Age- Drift* y *Cohort- Drift* presentan un razonamiento equivalente al que se acaba de realizar.

Por otro lado, los submodelos *A*, *P* y *C* incluyen restricciones tanto sobre las dobles diferencias como, sobre el plano lineal. En este aspecto, por ejemplo, el modelo *C* presentaría la siguiente estructura (para un *age- cohort array*):

$$\mu_{i,k} = \gamma_k$$

De tal forma, en este último modelo, se han eliminado las dobles diferencias correspondientes a las variables edad y periodo. Además, en relación con el plano lineal; este posee, únicamente, la pendiente correspondiente a la variable cohorte (las pendientes referentes a los regresores edad y periodo han sido restringidas).

Finalmente, en el modelo correspondiente al plano lineal (*linear plane model*) no existirían efectos no lineales; ya que, las dobles diferencias, habrían sido restringidas.

3.5.2.3. Submodelos surgidos a partir de distintas formas funcionales

La adopción de distintas formas funcionales, en relación con los efectos temporales de los modelos edad- periodo- cohorte, puede generar un nuevo conjunto de submodelos. Estos son especialmente utilizados en la práctica.

Por ejemplo, en algunas ocasiones, los efectos asociados a la variable edad poseen una apariencia, desde un punto de vista gráfico, de curva (pudiendo ser, en este aspecto, convexa o cóncava)¹². En estos casos, el regresor edad, puede ser definido a través de un polinomio de orden 2; teniendo asociado, pues, la siguiente hipótesis:

$$\alpha_i = \alpha_c + \alpha_l \times i + \alpha_q \times i^2 \quad (3.46)$$

De tal forma, este regresor poseería dobles diferencias constantes. Por otro lado, dicha restricción cuadrática podría ser introducida a través del parámetro canónico, poseyendo $A - 3$ grados de libertad.

Finalmente, como se señala en Fannon y Nielsen (2018) [91], aplicar una restricción a los efectos temporales del modelo, con el objetivo de convertirlos en un polinomio de orden k , es equivalente a aplicar una restricción, en relación con las dobles diferencias, con el objetivo de que estas se transformen en un polinomio de orden $k - 2$.

Por otro lado, también se podrían considerar modelos cuadráticos más complicados; en los que, los tres regresores temporales, poseyeran una forma cuadrática. El predictor de este hipotético modelo sería el siguiente (para un *age-cohort array*):

$$\mu_{ij} = \{\alpha_c + \alpha_l \times i + \alpha_q \times (i)^2\} + \{\beta_c + \beta_l \times j + \beta_q \times (j)^2\} + \{\gamma_c + \gamma_l \times k + \gamma_q \times (k)^2\} + \delta \quad (3.47)$$

siendo α_q , β_q y γ_q los parámetros no lineales del modelo. Estos elementos son, por tanto, identificables. Por otro lado, sobre este modelo podrían adoptarse restricciones que deriven, a su vez, en nuevos submodelos.

3.5.3. Extensiones de los modelos edad- periodo- cohorte

3.5.3.1. Modelos edad- periodo- cohorte mixtos

En los modelos edad- periodo- cohorte mixtos, algunos regresores son considerados aleatorios. En este aspecto, considerando un modelo cuya expresión de la variable cohorte esté sujeta a aleatoriedad, y teniendo en cuenta que, las variables periodo y edad poseen efectos fijos, se puede reestructurar el modelo general lineal de la siguiente forma:

$$y_{ijk} = \lambda + \alpha_i + \beta_j + u_k + \epsilon_{ijk} \quad (3.48)$$

donde y_{ij} es la variable respuesta para el i -ésimo grupo de edad, en el j -ésimo periodo y en la k -ésima cohorte. Por otro lado, λ es la constante del modelo; mientras que, α_i se corresponde con el efecto fijo referente al grupo de edad i . Igualmente, β_j es el efecto fijo referente al periodo j ; mientras que, u_k es el efecto aleatorio relativo a la cohorte número k . Finalmente, ϵ_{ijk} se corresponde con el residuo aleatorio del modelo. Se debe tener en cuenta que, una de las categorías relativas a las variables edad y periodo debe ser fijada como referencia; ya que, se ha considerado que los efectos de dichos regresores son fijos.

Por otro lado, no sería necesario añadir más restricciones al modelo; ya que, estas se encuentran incluidas, internamente, en el mismo. Por tanto, este puede ser identificado. Además, el modelo contenido en la ecuación 3.48 puede ser ampliado a través de la adición de características referentes a la variable cohorte¹³; pudiendo ser representado, pues, de la siguiente manera:

$$y_{ijk} = \lambda + \alpha_i + \beta_j + b_k \times \text{Característica de la cohorte} + u_k + \epsilon_{ijk} \quad (3.49)$$

En relación con este último modelo, si las variables adicionales referentes a las cohortes son suficientemente representativas, este hecho derivará en la reducción de la varianza asociada a los efectos aleatorios de

¹²Esto ocurre, por poner un ejemplo, al analizar la tasa de desempleo de una región; ya que, los mayores ratios de desempleo se acumulan en los grupos de personas con mayor y menor edad. Por tanto, la tasa de desempleo posee, en relación con la variable edad, y desde un punto de vista gráfico, forma de U.

¹³La adición de estas variables permitirá estudiar, en una mayor profundidad, las diferencias existentes entre las distintas cohortes.

la variable cohorte (u_k).

Por otro lado, en los modelos edad- periodo- cohorte mixtos, los efectos de las variables independientes pueden ser identificados. En este aspecto, los residuos del modelo, ϵ_{ijk} , y los efectos aleatorios correspondientes a la variable cohorte (en este caso), u_k , se encuentran distribuidos de forma independiente. Además, los efectos aleatorios y las variables independientes presentan una correlación igual a cero (Snijders, 2005) [121]. Por tanto, se obtendrían resultados similares a los arrojados por el modelo edad- periodo- cohorte ANOVA; ya que, los efectos aleatorios del regresor cohorte no estarían relacionados con los efectos de las variables edad y periodo. Por tanto, los efectos lineales de ambos grupos de regresores serían independientes (pudiendo ser identificados).

Si se considera, desde un punto de vista matricial, el modelo lineal mixto se podría demostrar, nuevamente, que este es identificable. En este aspecto, el modelo lineal mixto poseería la siguiente expresión:

$$y = X\beta + Zu + \epsilon \quad (3.50)$$

siendo y el vector que agrupa los valores de la variable respuesta y X la matriz de diseño referente a los efectos fijos. Por otro lado, β es un vector formado por los coeficientes de los efectos fijos del modelo; mientras que, Z es la matriz de diseño que posee los efectos aleatorios de, en este caso, la variable cohorte. Por tanto, u es un vector columna formado por los coeficientes de los efectos aleatorios del modelo. Finalmente, ϵ se corresponde con el vector formado por los residuos. Se debe tener en cuenta que, la elección relativa al factor que se utilizará como efecto aleatorio, incidirá en los valores estimados de los efectos fijos; por lo que, la elección de dicho regresor, revista de una notable importancia. En este aspecto, la variable tratada como aleatoria no posee, por su propia definición, una tendencia lineal; por lo que, los regresores considerados como fijos, se adueñarán de la tendencia lineal asociada a esta. Por tanto, los coeficientes asociados a las variables consideradas como fijas, no pueden ser tratados como los parámetros que generan la variable de interés; ya que, intrínsecamente, estos poseen incorporada la tendencia lineal de la variable aleatoria.

Por otra parte, a la hora de usar estos modelos, se debe tener un pensamiento fundado en relación con el comportamiento del regresor considerado como aleatorio; ya que, este último, debe ser una variable que no posea una tendencia fija (en comparación con el resto de las variables independientes, que poseen una tendencia clara en cuanto a la generación de la variable respuesta). En este aspecto, podría darse el caso en el que, más de una de las variables independientes, posea efecto aleatorio; pudiendo incluso, ser considerados, todos los regresores, como variables aleatorias. En este hipotético caso, por tanto, no existirían efectos fijos. De tal forma, dichos modelos podrían ser identificados sin necesidad de aplicar ninguna restricción; aunque, no se solucionaría el problema relacionado con la manera en la que se distribuye la tendencia, en relación con el resto de variables independientes.

3.5.3.2. Modelos edad- periodo- cohorte jerárquicos

En este epígrafe, se hará referencia al procedimiento de resolución de modelos edad- periodo- cohorte desarrollado por Yang y Land (2006) [120]. Este método es utilizado en aquellas ocasiones en las que los datos posean un nivel individual; siendo este el caso, por ejemplo, de las encuestas de carácter transversal. Para analizar datos que poseen dichas características, estos autores recomiendan utilizar un modelo jerárquico en el que, la primera de las capas, esté formada por los niveles de la variable edad reflejados, estos últimos, en términos individuales. Por otro lado, los regresores periodo y cohorte deben estar representados, a nivel agregado, en la segunda capa. De tal forma, este tipo de modelos posibilita la inclusión de otros regresores ajenos a las variables de índole temporal; siendo, esta, una de las principales ventajas asociadas a este método.

Teniendo en cuenta la notación de los modelos jerárquicos, la primera capa de estos poseería la siguiente expresión:

$$Y_{ijk} = \beta_{0jk} + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + \dots + \beta_p X_{Pijk} + e_{ijk} \quad (3.51)$$

siendo $i = 1, 2, \dots, n_{jk}$. Además, j refleja el periodo ($j = 1, \dots, J$) y k representa la cohorte ($k = 1, \dots, K$). Por otra parte, $e_{ij} \sim N(0, \sigma^2)$.

Por otro lado, la segunda capa del modelo posee la siguiente expresión; caracterizándose, esta última, por modelar el comportamiento entre celdas:

$$\beta_{0jk} = \lambda_0 + u_{0j} + v_{0k} \quad (3.52)$$

siendo $u_{0j} \sim N(0, \tau_u)$ los efectos aleatorios del modelo; mientras que, $v_{0k} \sim N(0, \tau_v)$ reflejan los efectos aleatorios correspondientes a la variable cohorte. Por tanto, esta última expresión modela la constante de la ecuación 3.51.

Combinando las expresiones introducidas anteriormente, se obtiene el modelo de efectos mixtos que se muestra a continuación:

$$Y_{ijk} = \lambda_0 + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + \dots + \beta_p X_{Pijk} + u_{0j} + v_{0k} + e_{ijk} \quad (3.53)$$

siendo λ_0 el valor esperado de la variable dependiente cuando todos los regresores poseen un valor igual a cero. Por otro lado, u_{0j} se corresponde con el residuo asociado al periodo j ; poseyendo, este último, un comportamiento aleatorio; mientras que, v_{0k} es el residuo aleatorio correspondiente a la k -ésima cohorte.

Por otra parte, estos modelos pueden ser identificados, como señalan autores como Yang et al. (2008) [118]. En este aspecto, considerando un parámetro no lineal relacionado con una de las variables independientes del modelo (edad, periodo o cohorte), se resolvería el problema de identificación asociado a los modelos edad- periodo- cohorte. Generalmente, se suele considerar la variable edad al cuadrado; en relación, esta última, con los efectos fijos del modelo. Por otro lado, según dichos autores, también se podría identificar el modelo a través del uso de intervalos desiguales relacionados con la medición de las variables edad, periodo y cohorte; ya que, a través de este proceso, se evitaría que existiese una dependencia lineal exacta entre dichos regresores. Sin embargo, el uso de este procedimiento está reservado, únicamente, para aquellos conjuntos de datos construidos de manera individual. Además, se debe tener en cuenta que, los resultados obtenidos mediante el uso de un modelo jerárquico aplicado a este tipo de datos, dependen de la amplitud asociada a cada intervalo (Luo y Hodges, 2013) [122]. A pesar de este hecho, dicho procedimiento es especialmente utilizado en la práctica. Por otra parte, como también señalan Yang y Land (2013) [66] el modelo edad- periodo- cohorte jerárquico no se encuentra sometido al problema de identificación clásico asociado a los modelos edad- periodo- cohorte; ya que, los tres regresores temporales del modelo, no presentan las propiedades de linealidad y aditividad en la misma capa.

Por otro lado, el modelo edad- periodo- cohorte jerárquico, presenta importantes similitudes en relación con el modelo edad- periodo- cohorte mixto; aunque, este último utiliza, únicamente, información de carácter agregado. En este aspecto, estos modelos no presentan problemas de identificación; aunque, generan estimaciones sesgadas de los parámetros. De tal forma, el valor estimado asociado a los efectos fijos depende de la variable que sea supuesta como aleatoria.

3.6. Cuestiones que pueden, o no pueden, resolverse

Como se ha señalado a lo largo de estos epígrafes, los modelos edad- periodo- cohorte no resuelven el problema de identificación asociado a los regresores del modelo; por lo que, las partes lineales de estos modelos, no pueden ser identificadas. Este hecho, limita las cuestiones que podrán ser resueltas a través de la implementación de este tipo de modelos.

En primer lugar, respecto a las preguntas que pueden ser abordadas, destacan las siguientes:

- Preguntas relacionadas con el procedimiento de *difference- in- difference*. Dichas cuestiones pueden ser analizadas utilizando modelos edad- periodo- cohorte.
- Preguntas relacionadas con los efectos no lineales referentes a las variables edad, periodo y cohorte. Este tipo de análisis es utilizado, principalmente, en investigaciones de naturaleza económica, como señalan autores como Fannon y Nielsen (2018) [91].

- Análisis de carácter exploratorio, en relación con las variables edad, periodo y cohorte.
- Realización de predicciones. De tal forma, considerando este tipo de modelos, la extrapolación es necesaria al realizar actividades de índole predictiva. Por otro lado, si el método de extrapolación escogido es invariante ante el problema de identificación del modelo, se obtendrán estimaciones robustas; siendo, estas, independientes en relación con las suposiciones adoptadas (Kuang et al., 2008) [5].
- Preguntas que no poseen relación con los efectos temporales. De tal forma, se pueden utilizar los modelos edad- periodo- cohorte como elementos de control.

Por otro lado, en cuanto a las cuestiones que no pueden ser abordadas utilizando los modelos edad- periodo- cohorte, destacan aquellas referentes a las partes lineales del modelo. En este aspecto, la adopción de restricciones no resolvería este problema, como se señaló anteriormente. De tal forma, lo recomendable sería formular hipótesis relacionadas con la parte del modelo no lineal; ya que, si se formularan hipótesis relacionadas con la parte lineal, se obtendrían resultados sesgados.

3.7. Alternativas a los modelos edad- periodo- cohorte

Los problemas de identificación asociados a los modelos edad- periodo- cohorte han provocado el surgimiento de otras alternativas que abogan por abandonar este tipo de modelos. Se señalarán, en este aspecto, dos de dichas alternativas:

- Análisis gráfico.
- Aproximación factor- característica.

Estos procedimientos serán descritos en los siguientes epígrafes.

3.7.1. Análisis gráfico

Autores como Meghir y Whitehouse (1996) [108] o Voas y Chaves (2016) [109] utilizan el análisis gráfico con el objetivo de identificar las variables temporales que modelan el comportamiento de la variable dependiente; evitando, de esta manera, el problema de identificación asociado a los modelos edad- periodo- cohorte. De hecho, algunos autores señalan que, debido al problema de identificación, no existiría un interés práctico en ir más allá del análisis gráfico. Sin embargo, en la mayor parte de las ocasiones, no se pueden extraer conclusiones a través de este método, ante la inexistencia de paralelismo en las tendencias observadas. Además, esta metodología no permite la ejecución de análisis formales; así como, de otras herramientas de verificación de modelos. En base a estas circunstancias, lo más adecuado consistiría en incorporar estos métodos gráficos a la hora de ejecutar un modelo edad- periodo- cohorte; sin descuidar, a su vez, la parte analítica. De hecho, los métodos gráficos pueden poseer una notable utilidad a la hora de determinar las variables relevantes que explican el comportamiento de la variable dependiente.

3.7.2. Aproximación factor- característica

En el epígrafe 3.5 se describieron técnicas que posibilitaban el cálculo de la varianza asociada, de forma individual, a cada uno de los regresores temporales del modelo (edad, periodo y cohorte). Por tanto, a la hora de trabajar con la extensión ANOVA de los modelos edad- periodo- cohorte, o con los modelos edad- periodo- cohorte mixtos, no se deben considerar los efectos fijos, arrojados por dichos modelos, como estimaciones insesgadas de aquellos parámetros que generan los valores de la variable respuesta.

Partiendo de esta idea, se procederá a introducir una estrategia que posibilita la ruptura de la linealidad existente entre las variables edad, periodo y cohorte; pudiéndose estimar, de esta manera, los efectos asociados a dichas variables. Para tal fin, se deben sustituir uno, o varios, de dichos regresores por una

característica referente a los grupos de dichas variables.

En cuanto a las ventajas asociadas a esta metodología, destaca el hecho de que, esta última, genera un procedimiento que permite que las variables edad, periodo y cohorte incidan en el valor de la variable respuesta. Por otro lado, uno de los principales inconvenientes asociados a esta metodología consiste en que, el factor que representa a la variable independiente tradicional (edad, periodo o cohorte), no permite recoger, completamente, el efecto total de la variable que sustituye; por lo que, el efecto provocado por el regresor al que este representa, se verá subestimado. De esta manera, debido a las características asociadas a los modelos edad- periodo- cohorte, el resto de variables independientes temporales se adueñarán de los efectos lineales del regresor que ha sido sustituido por otro factor, con los efectos distorsionadores que esto provocará sobre el modelo.

A continuación, se introducirán los casos más tradicionales de uso de la aproximación factor- característica; describiendo, a su vez, las cuestiones más relevantes asociadas a su implementación.

3.7.2.1. Aproximación por características usando un factor

Desde un punto de vista tradicional, se suele hacer uso de una única característica; en relación, a su vez, con un único factor. En este aspecto, normalmente, se suele usar un factor en lugar de la variable cohorte; aunque, autores como O'Brien et al. (1999) [123] utilizan dos características, en relación con este regresor. Sin embargo, Winship y Harding (2008) [124], utilizan distintos factores para múltiples variables.

Con el objetivo de estudiar las características asociadas a estos modelos; se introducirá, a continuación, un modelo edad- periodo- cohorte con dos características que reflejan a la variable cohorte:

$$Y_{ij} = \lambda + \alpha_i + \beta_j + \chi_1 + \chi_2 + \epsilon_{ij} \quad (3.54)$$

donde Y_{ij} refleja el valor de la variable salida para el grupo de edad i y el periodo j . Además, λ es la constante; mientras que, α_i y β_j representan, respectivamente, el i -ésimo grupo de edad y el j -ésimo periodo. Por otro lado, χ_1 y χ_2 reflejan los coeficientes asociados a las dos características de la variable cohorte. Finalmente, el término de error viene determinado por el elemento ϵ_{ij} .

Desde un punto de vista matricial, esta última ecuación puede reescribirse de la siguiente manera:

$$y = Xb + \epsilon \quad (3.55)$$

siendo y un vector de rango $I \times J$ que agrupa los valores de la variable respuesta, X la matriz de diseño y ϵ un vector de orden $I \times J$ que refleja los residuos del modelo. En este caso, la matriz de diseño no presenta dependencia lineal; por lo que, no existe ningún problema de identificación asociado a este modelo. Sin embargo, al sustituir con las dos variables de características al regresor cohorte, los efectos lineales de este último factor son captados por el resto de variables independientes (en este caso, por los regresores edad y periodo). Por tanto, se obtendrán estimadores sesgados relacionados con los efectos de las variables edad, periodo y cohorte.

Por otro lado, la relación que existe entre los efectos de las variables edad, periodo y cohorte puede ser expresada de la siguiente manera:

$$\begin{aligned} t_\alpha^* &= t_\alpha + k \\ t_\beta^* &= t_\beta - k \\ t_\gamma^* &= t_\gamma + k \end{aligned} \quad (3.56)$$

En este caso, t_α , t_β y t_γ representan las tendencias lineales de las variables edad, periodo y cohorte, respectivamente, considerando una restricción. Por otro lado, t_α^* , t_β^* y t_γ^* reflejan las tendencias lineales asociadas a otra restricción. De tal forma, en base a esta relación, un cambio en la restricción (representado por el parámetro k en la anterior relación) produce una variación en la pendiente del resto de los coeficientes del modelo. Este hecho debe tenerse en consideración a la hora de interpretar dichos modelos.

Por otro lado, si se excluye una de las variables temporales a la hora de estimar el modelo, los regresores que continúan en el mismo absorberían la tendencia lineal correspondiente a la variable eliminada. De tal forma, esta situación queda reflejada en el sistema introducido en la expresión 3.56. Sin embargo, cuando se incorporan una o dos características al modelo, con el objetivo de sustituir a un factor temporal, las consecuencias son ligeramente distintas; ya que, esto no provocaría que la tendencia de dicho factor tomase un valor igual a cero. En este aspecto, si las características permitiesen reflejar, de forma bastante aproximada, la tendencia inherente a la variable cohorte, el resto de regresores no capturarían la tendencia lineal asociada a las cohortes. Sin embargo, este hecho podría comprobarse, únicamente, a través de información externa.

3.7.2.2. Aproximación por características usando dos o más factores

En base a las aportaciones de autores como O' Brien (2014) [104], los modelos con características para dos o más factores no son frecuentes en la literatura de los modelos edad- periodo- cohorte. En este aspecto, como bien señala este autor, destacaría el trabajo de Winship y Harding (2008) [124], que utilizan un modelo de ecuaciones estructurales, con múltiples características, con el objetivo de modelar una serie de factores.

Desde el punto de vista de las ventajas asociadas a este tipo de modelos, destaca el hecho de que este estaría identificado. Además, las relaciones estadísticas que provocan que, las variables sustituidas por características sean relevantes, serían mostradas de forma clara; pudiéndose medir, de manera directa, los efectos de las mismas. Sin embargo, si las características no tuviesen una relación estrecha con las variables que estas sustituyen, la tendencia lineal de dichos regresores quedaría absorbida por la variable restante, generando una estimación sesgada relacionada con los efectos de esta última.

3.7.2.3. Comparación de la descomposición de la varianza

Al igual que se dijo con anterioridad, cuando se consideran dos variables en el modelo; siendo, la tercera de ellas, representada por una característica; las dos variables restantes absorberían la varianza que no pudiera ser representada por dicha característica. Partiendo de esta idea, se pretende describir el procedimiento de cálculo de esta proporción de varianza.

Para ello, en primer lugar, se debe estimar el modelo con las variables edad, periodo y cohorte; escogiendo, para tal fin, una restricción que permita resolver el problema de identificación asociado a este tipo de modelos. De tal forma, independientemente de la restricción adoptada, se obtendrá un R^2 determinado. Tras esto, se estima el modelo, utilizando, únicamente, dos regresores. De tal forma, la diferencia existente entre el R^2 anterior y el R^2 obtenido a través de esta última estimación, representa la proporción de varianza de la variable dependiente explicada por el regresor que ha sido excluido del modelo. Posteriormente, se vuelve a estimar el modelo; aunque, en esta ocasión, se deben considerar los dos factores temporales, además de, una característica que represente al tercer factor. Por tanto, considerando que se ha incorporado una característica de la variable cohorte, el porcentaje de varianza asociado a dicho regresor que es cuantificado, a su vez, por la característica que sustituye a esta variable, puede obtenerse a partir de la siguiente expresión (O' Brien, 2014; pp. 157) [104]:

$$\left[\frac{R_{apcc}^2 - R_{ap}^2}{R_{apc}^2 - R_{ap}^2} \right] \times 100 \quad (3.57)$$

donde las siglas *apc* se corresponden con el modelo que contiene a las variables edad, periodo y cohorte; mientras que, *ap* hace alusión al modelo que contiene, únicamente, a las variables edad y periodo. Finalmente, las siglas *apcc* representan al modelo que posee la característica que sustituye a la variable cohorte; además de, a los regresores edad y periodo.

Capítulo 4

Paquete *apc*

Aunque existen otros paquetes en R¹ y Stata² que permiten implementar modelos de tipología edad-periodo-cohorte; se utilizará, en este estudio, el paquete *apc*³ (Fannon y Nielsen, 2020) [8], que se encuentra disponible en el software R. Para consultar información acerca de este paquete se puede ejecutar, en la consola de R, la instrucción `help (apc)`. Por otro lado, en Nielsen (2015) [129] y Fannon y Nielsen (2020) [8] se puede consultar información acerca de este paquete; además de, algunos ejemplos que permiten ilustrar la utilización del mismo.

Por otra parte, el paquete *apc* utiliza la parametrización canónica propuesta por Kuang et al. (2008) [5], que fue descrita en el apartado 3.3.1 del presente estudio. En términos generales, esta metodología divide al predictor en una parte lineal (no identificable) y una parte no lineal (que puede ser identificada). De tal forma, el predictor queda conformado por las dobles sumas de las segundas diferencias de los regresores edad, periodo y cohorte; además de, por tres parámetros reflejados en función de un plano lineal, siendo, estos, el nivel y dos pendientes.

Partiendo de estas ideas, en este epígrafe se analizará el paquete *apc*, con el objetivo de realizar, en el siguiente apartado del proyecto, una aplicación práctica del mismo. Para estudiar dicho paquete, se seguirán los siguientes pasos:

1. Creación de un objeto de datos que permite definir la base de datos a estudiar.
2. Realización de actividades de visualización de los datos; así como, de exploración de los mismos.
3. Estimación del modelo edad- periodo- cohorte y de los distintos submodelos; empleando, para tal fin, la parametrización canónica propuesta por Kuang et al. (2008) [5].
4. Realización de actividades de estimación e inferencia sobre los modelos obtenidos en el paso anterior.

De tal forma, para ilustrar estos pasos, se utilizará uno de los conjuntos de datos que incluye el paquete *apc* (ver Fannon y Nielsen, 2020; pp. 67- 89 [8]); siendo estos, los siguientes:

- **Data.aids.** Esta base de datos ha sido extraída de De Angelis y Gilks (1994) [130] y representa el retraso en el diagnóstico del SIDA (*AIDS*), concerniente a la población de Reino Unido. De tal forma, en la base de datos, se tiene en cuenta el número de casos y el tiempo de demora asociado al diagnóstico de la enfermedad. Además, los datos son trimestrales; presentando, a su vez, una estructura trapezoidal.
- **Data.asbestos.** Extraída de Martínez- Miranda et al. (2014) [87], representa el número de muertes causadas, en Reino Unido, por el cáncer mesotelioma. Esta enfermedad está asociada con la exposición a los asbestos. Por otro lado, la base de datos presenta una matriz de datos en formato *PA*.

¹Destaca, por ejemplo, el paquete *Epi* (Carstensen et al., 2021) [125].

²Destacan, por ejemplo, los paquetes *apc* (Schulhofer- Wohl y Yang, 2006) [126], *st0245* (Sasieni, 2012) [127] y *apcd* (Chauvel, 2012) [128].

³En este aspecto, el paquete *apc* permite aplicar modelos edad- periodo- cohorte a tres tipos de bases de datos (*data- array*): bases de datos de carácter agregado, datos de corte transversal recogidos de forma repetida y datos en formato de panel.

- **Data.Belgian.lung.cancer.** Representa los ratios anuales de incidencia, por 100000 habitantes, del cáncer de pulmón en mujeres belgas. El periodo de análisis se encuentra contenido entre los años 1955 y 1978; poseyendo, la base de datos, estructura de *AP* array. Estos datos fueron extraídos de la *WHO* (*World Health Organization*) y están contenidos, a su vez, en Clayton y Schiffers (1987) [131].
- **Data.Italian.bladder.cancer.** Esta base de datos ha sido extraída de Clayton y Schiffers (1987) [131]; contenido los ratios de incidencia, por 100000 habitantes, de cáncer de vejiga en hombres italianos. Además, el periodo de análisis abarca desde el año 1955 hasta 1979; mientras que, la base de datos, posee un formado edad- periodo (*age- period array*).
- **Data.Japanese.breast.cancer.** Esta base de datos también ha sido extraída de Clayton y Schiffers (1987) [131]; conteniendo, a su vez, los ratios de mortalidad (por 100000 habitantes) de cáncer de mama existentes Japón (periodo 1955- 1979). La base de datos presenta una estructura de edad- periodo, con grupos de edades y periodos de 5 unidades (Units=5).
- **Data.RH.mortality.** Esta base de datos ha sido extraída de Jacobsen et al. (2004) [132] y Riebler y Held (2010) [133], representando los datos de mortalidad de mujeres danesas y noruegas. Por otro lado, la base de datos posee un formato edad- periodo (*Age- period array*).
- **Data.US.prostate.cancer.** Esta base de datos ha sido extraída de Holford (1983) [100], representando las muertes específicas causadas, en EE.UU, por el cáncer de próstata. En concreto, abarca los periodos acaecidos entre 1935 y 1969, considerando intervalos de 5 años. Además, la base de datos se encuentra en formato edad- periodo (*age- period array*).

En este caso, se usará la base de datos **Data.Italian.blader.cancer** como ejemplo para proceder a explicar las características del paquete *apc*. Para ello, se seguirán los pasos que se introdujeron anteriormente, combinando salidas de resultados obtenidos a través del software R⁴ y explicaciones plasmadas en el software L^AT_EX.

4.1. Creación del objeto de datos

Lo primero que se debe realizar es introducir la instrucción *library(apc)*, con el objetivo de cargar el paquete *apc*. Por otro lado, en este contexto, el análisis de un modelo edad- periodo- cohorte puede ser realizado considerando dos situaciones:

- Un caso en el que tanto las dosis (exposición, casos, ...) como las respuestas (número de muertes, por ejemplo) estén disponibles.
- Una situación en la que se disponga, únicamente, de la matriz de respuestas.

De tal forma, si se quieren analizar, directamente, los ratios de mortalidad se deberían escoger, estos datos, como matriz de respuestas. Esto será explicado posteriormente.

Teniendo en cuenta estas consideraciones, y tras cargar el paquete *apc*, se debe organizar la base de datos que se va a utilizar, mediante la creación de un objeto que permita implementar el análisis computacional. De tal forma, estas acciones pueden ser realizadas a través de la función *apc.data.list*, que posee los argumentos que se muestran a continuación:

- **response.** Hace alusión a la matriz, o vector, con el número de respuestas. Tiene que tener un formato equivalente al argumento **data.format**, que se muestra a continuación. En este aspecto, los argumentos **response** y **data.format** son los dos únicos argumentos obligatorios de esta función.
- **data.format.** Este argumento es de tipo *character*, poseyendo distintas opciones de elección, en función de las filas (primera letra) y de las columnas (segunda letra) de la matriz de datos:
 - *AC.* Representa una base de datos de tipo edad- cohorte.

⁴En concreto, se ha utilizado el IDE R- Studio.

- *AP*. Representa una base de datos de tipo edad- periodo.
 - *CA*. Representa una base de datos de tipo cohorte- edad.
 - *CL*. Representa una base de datos de tipo cohorte- edad; que presenta, a su vez, forma triangular.
 - *CP*. Representa una base de datos de tipo cohorte-periodo.
 - *PA*. Representa una base de datos de tipo periodo- edad.
 - *PC*. Representa una base de datos de tipo periodo- cohorte.
 - *trapezoid*. Representa una base de datos de tipo edad- periodo, donde la diagonal de la variable periodo posee el valor *NA*.
- **doses**. Hace referencia a una matriz con el número de dosis. Debe tener el mismo formato que el argumento **response**.
 - **age1**. Etiqueta temporal para el grupo de edad más joven. Se puede usar en aquellos casos en los que la variable edad aparece en la base de datos. Toma un valor, por defecto, igual a 1.
 - **per1**. Etiqueta temporal para el grupo de periodos más antiguo. Por tanto, este argumento puede utilizarse, únicamente, en aquellos casos en los que la variable periodo sea incluida en la base de datos. Por defecto, tiene un valor igual 1.
 - **coh1**. Etiqueta temporal para el grupo de edad más joven. Este argumento puede utilizarse en aquellos casos en los que, la variable cohorte, sea incluida en la base de datos Toma un valor, por defecto, igual a 1.
 - **unit**. Este argumento se utiliza con el objetivo de distribuir, de manera uniforme, las medidas de edad, periodo y cohorte. Por ejemplo, si los datos son mensuales, $Unit = \frac{1}{12}$; mientras que, por defecto, toma un valor igual a 1.
 - **per.zero**. Se utiliza en aquellos casos en los que la base de datos es trapezoidal.
 - **per.max**. Se utiliza en aquellos casos en los que la base de datos es trapezoidal.
 - **time.adjust**. Las etiquetas temporales se basaban en los argumentos **age1**, **per1** y **coh1**. La etiqueta temporal restante se obtiene a partir de la fórmula $age1 + coh1 = per1 + time.adjust$. Toma un valor, por defecto, igual a 0.
 - **label**. Posee una tipología tipo *string*; permitiendo dar una etiqueta al objeto creado.

En el caso actual, en el que se estudiará la base de datos **Data.Italian.bladder.cancer**, se cargarán los datos de forma directa; ya que, los argumentos de la función *apc.data.list* que se acaban de introducir, vienen integrados en el paquete *apc* de forma automática. Por tanto, en base a estas ideas, se ejecutarán las siguientes instrucciones:

```
> library(apc)
> base_datos=data.Italian.bladder.cancer
```

obteniéndose las siguientes salidas:

```
> base_datos
## $response
##      1955-1959 1960-1964 1965-1969 1970-1974 1975-1979
## 25-29         3         3         1         4         12
## 30-34        16        17        11         8         8
## 35-39        24        29        33        39        30
## 40-44        79        76        82        95       115
## 45-49       234       185       183       267       285
## 50-54       458       552       450       431       723
## 55-59       720       867      1069       974      1004
## 60-64       890      1230      1550      1840      1811
```

```

## 65-69      891      1266      1829      2395      3028
## 70-74      920      1243      1584      2292      3176
## 75-79      831       937      1285      1787      2659
##
## $dose
##      1955-1959 1960-1964 1965-1969 1970-1974 1975-1979
## 25-29 100.00000 100.00000 100.00000 100.00000 100.00000
## 30-34  94.11765  94.44444  91.66667 100.00000  88.88889
## 35-39  75.00000  93.54839  94.28571  92.85714  93.75000
## 40-44  75.96154  72.38095  90.10989  91.34615  90.55118
## 45-49  81.81818  73.41270  70.11494  87.82895  90.18987
## 50-54  68.97590  78.52063  69.98445  66.71827  85.36009
## 55-59  56.64831  64.74981  73.26936  66.53005  61.29426
## 60-64  44.25659  51.29274  58.07419  66.78766  63.47704
## 65-69  36.51639  38.17853  43.42355  50.13607  60.11515
## 70-74  28.04023  29.37840  29.96028  34.72201  42.55091
## 75-79  18.24769  19.54526  20.70911  21.11045  25.51579
##
## $data.format
## [1] "AP"
##
## $age1
## [1] 25
##
## $per1
## [1] 1955
##
## $coh1
## [1] 5
##
## $unit
## [1] 5
##
## $per.zero
## NULL
##
## $per.max
## NULL
##
## $time.adjust
## [1] 0
##
## $label
## [1] "Italian bladder cancer"
##
## $n.decimal
## NULL

```

En base a estas salidas, se tiene que la base de datos posee un formato *AP*, comenzando en 1955 y finalizando en 1979. Además, como $Unit = 5$, los grupos de las variables edad y periodo se agrupan en intervalos de 5 años. En este aspecto, el argumento *base_datos\$response*, hace alusión a la matriz de casos (*matrix of cases*); mientras que, el argumento *base_datos\$dose*, hace referencia a la matriz de casos/ratios (*matrix cases/rates*).

4.2. Exploración de los datos

Una vez introducida la base de datos, se procederá con la representación de la misma. En este aspecto, el paquete *apc* permite representar, a través de la función *apc.plot.data.all*, una importante cantidad de gráficos. De tal forma, esta función posee los siguientes argumentos:

- **apc.data.list**. Este argumento fue definido en el epígrafe anterior, reflejando el objeto de datos con el que se trabajará. Además, este es el único argumento obligatorio que posee esta función.
- **log**. Este se corresponde con un argumento de tipo carácter; siendo, además, opcional. De tal forma, toma el valor “y” si la escala es logarítmica; mientras que, en cualquier otro caso, toma el valor “”. Por defecto, este argumento toma un valor igual a “y”.
- **rotate**. Este argumento de la función presenta una tipología lógica. Si se le otorga un valor igual a TRUE, las gráficas giran, noventa grados, en el sentido de las agujas del reloj. Si el formato de los datos es de tipo *CL* (**data.format=CL**), los gráficos girarán, noventa grados, en el sentido contrario al de las agujas del reloj. Este argumento adopta un valor lógico, por defecto, igual a FALSE.

Una vez introducida la descripción de la función, se procederá con la implementación de la misma. Para ello, se ejecutará el siguiente comando:

```
> apc.plot.data.all(base_datos)
```

En este caso aparecen, en la consola de R, los siguientes *warnings*:

```
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"  
  
## [1] "apc.plot.data.within warning: maximal index not divisible by thin, so last group  
smaller than other groups"
```

Si se consulta la información suministrada por Fannon y Nielsen (pp. 49-50; 2020) [8], en relación con la aparición de *warnings* al implementar esta función; se tiene que, dichas advertencias, están relacionadas con los datos elegidos. De tal forma, para solucionar dicha cuestión, se debe trabajar con los gráficos individuales. En este aspecto, la función *apc.plot.data.within* posee un rol destacado, como se ilustrará a continuación.

Partiendo de esta cuestión, se implementará, en primer lugar, la función *apc.data.sums*, que posibilita el cálculo de las sumas de cada uno de los regresores a nivel individual. Dicha función posee los siguientes argumentos:

- **apc.data.list.** El primer argumento de esta función hace referencia a la lista *apc.data.list* que fue descrita, a su vez, anteriormente.
- **data.type.** Este argumento de tipo carácter no es obligatorio. Además, posee las siguientes opciones de elección:
 - “r”. Se debe introducir este *string* cuando se desea computarizar las sumas para las respuestas (*responses*).
 - “d”. Esta opción debe ser escogida cuando se quieren calcular las sumas correspondientes a la matriz de dosis (*doses*).
 - “m”. Se debe escoger esta opción cuando se desean calcular las sumas relacionadas con las ratios de mortalidad (*mortality rates*). En este aspecto, se debe recordar que los ratios se calculan como el cociente entre las matrices de respuesta y dosis.
 - “a”. Esta opción debe ser escogida cuando se desea computarizar las sumas para todos los elementos citados anteriormente: (*responses*, *dose* y *mortality rates*). Esta es la elección que viene implementada por defecto.
- **average.** Este argumento opcional posee una naturaleza lógica. Si toma un valor TRUE, se calcularán las medias; mientras que, si toma un valor FALSE, se calculan las sumas. Este argumento toma, por defecto, un valor igual a FALSE.
- **apc.index.** Este argumento permite generar una lista. En caso de no especificarse, dicha lista se calcula de forma automática.
- **Otros argumentos.** Aparte de los argumentos que se acaban de señalar, la función *apc.data.sums* posee otros argumentos, tales como: el tipo de elemento que se utilizará en el gráfico (**type**), la escala que se empleará (**log**), el título principal del gráfico (**main.outer**) y el título que se asociará a los subgráficos (**main.sub**).

Por tanto, se aplicará la función *apc.data.sums* a la base de datos **data.Italian.bladder.cancer**, realizando las modificaciones pertinentes en el argumento **data.type**. De tal forma, se mostrarán, a continuación, las sumas para las respuestas, dosis, tasas de mortalidad y medias de las tasas de mortalidad:

```
> apc.data.sums(base_datos,data.type ="r") #responses
## $sums.age
## [1] 23 60 155 447 1154 2614 4634 7321 9409 9215 7499
##
## $sums.per
## [1] 5066 6405 8077 10132 12851
##
## $sums.coh
## [1] 831 1857 3419 5527 8730 8446 6723 3499 1718 1117 433 168 39 12 12

> apc.data.sums(base_datos,data.type ="d") #doses
## $sums.age
## [1] 500.0000 469.1176 449.4412 420.3497 403.3646 369.5593 322.4918 283.8882
## [9] 228.3697 164.6518 105.1283
```

```

##
## $sums.per
## [1] 679.5825 715.4519 741.5981 778.0368 801.6932
##
## $sums.coh
## [1] 18.24769 47.58550 86.60390 133.50585 211.60240 284.48688 360.51098
## [8] 349.36578 345.50842 450.96497 470.26619 375.07499 293.75000 188.88889
## [15] 100.00000

> apc.data.sums(base_datos,data.type ="m") #mortality rates

## $sums.age
## [1] 0.23 0.64 1.72 5.31 14.19 35.03 71.71 126.86 197.82 268.64
## [11] 344.39
##
## $sums.per
## [1] 146.63 171.90 208.75 251.70 287.56
##
## $sums.coh
## [1] 45.54 80.75 128.76 190.79 249.03 169.13 102.40 53.16 26.82 12.90
## [11] 4.76 1.84 0.41 0.13 0.12

> apc.data.sums(base_datos,data.type ="m", average="TRUE") #average mortality rates

## $sums.age
## [1] 0.046 0.128 0.344 1.062 2.838 7.006 14.342 25.372 39.564 53.728
## [11] 68.878
##
## $sums.per
## [1] 13.33000 15.62727 18.97727 22.88182 26.14182
##
## $sums.coh
## [1] 45.5400000 40.3750000 42.9200000 47.6975000 49.8060000 33.8260000
## [7] 20.4800000 10.6320000 5.3640000 2.5800000 0.9520000 0.4600000
## [13] 0.1366667 0.0650000 0.1200000

```

Para representar estas últimas matrices de sumas, se aplicará la función *apc.plot.data.sums*, que posee los siguientes argumentos:

- **apc.data.list**. Es el argumento que representa la lista que debe introducirse al comenzar con el análisis de un modelo edad- periodo- cohorte. Ha sido descrito anteriormente.
- **data.type**. Este argumento es opcional; siendo, además, de tipo carácter. En este aspecto, se pueden escoger las siguientes opciones:
 - “r”. Si se desean obtener los gráficos de las sumas asociadas a la matriz de respuestas (*responses*).
 - “d”. Cuando se quiere obtener los gráficos para las sumas de las dosis (*doses*).
 - “m”. Se debe utilizar este elemento cuando se quieren obtener las gráficas referentes a las sumas de los ratios de mortalidad.
 - “a”. Esta opción debe ser escogida cuando se quieren obtener las gráficas de las sumas pertenecientes a todos los casos anteriores.
- **average**. Argumento opcional de tipo lógico. Si se quieren obtener los gráficos de las medias, se debe escoger la opción TRUE; mientras que, si se desean obtener los gráficos de las sumas, se debe señalar la opción FALSE. Esta última es la que viene implementada por defecto.

- **type**. Este argumento opcional permite modificar las características de los puntos del gráfico. De tal forma, se puede escoger entre puntos y líneas superpuestas (“o”), líneas (“l”) o puntos (“p”). La opción “o”, es la que aparece implementada por defecto.
- **log**. Argumento opcional de tipo carácter. Si la variable de interés está asociada a una escala logarítmica, se debe escoger la opción “y”.
- **main.outer**. Argumento opcional de tipo carácter que permite dar un título principal al gráfico.
- **main.sub**. Argumento opcional de tipo carácter que posibilita la adición de títulos a los sub-gráficos.

Los resultados obtenidos a partir de la ejecución de dicha función se muestran en la figura 4.1.

```
> apc.plot.data.sums(base_datos,data.type ="a")
```

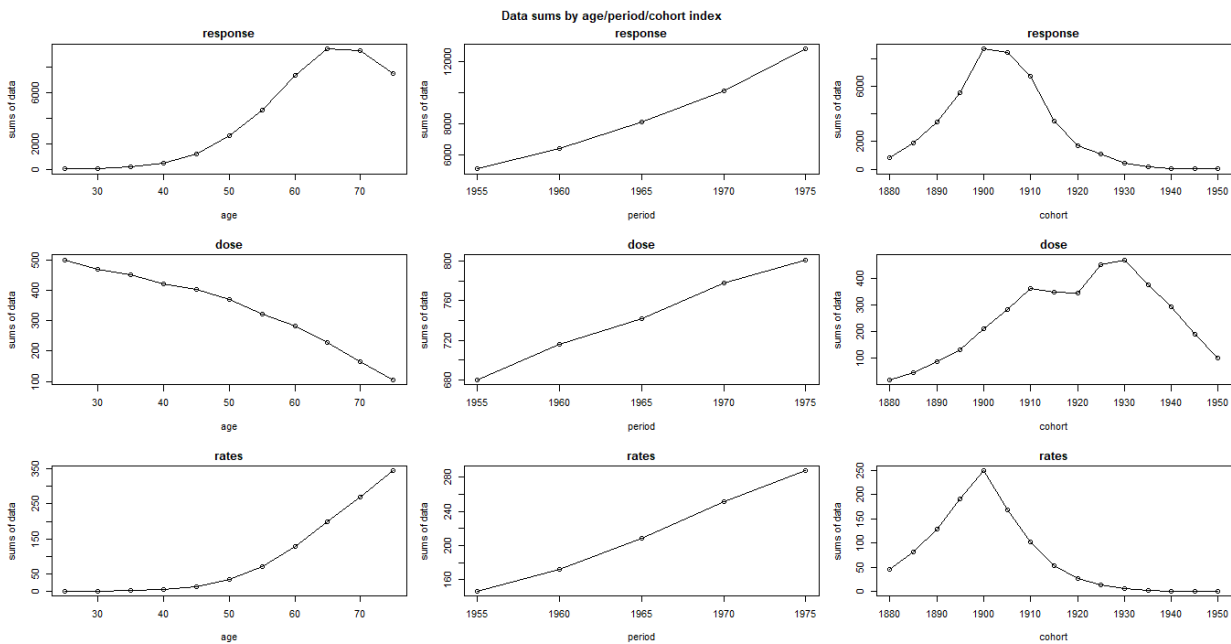


Figura 4.1: Resultados gráficos obtenidos al ejecutar la función *apc.plot.data.sums*.

Por otro lado, la función *apc.plot.data.sparsity*, posibilita la visualización de cifras con valores reducidos en la matriz de datos. Los argumentos de esta función son los siguientes:

- **apc.data.list**. Objeto de datos con el que se ha trabajado a lo largo del análisis.
- **data.type**. Argumento de tipo carácter opcional. Posee las siguientes opciones de elección: “r”, para las respuestas (*responses*), “d”, respecto a las dosis (*dose*), “m”, en relación con las tasas de mortalidad y finalmente “a”, para representar todas las opciones anteriores.
- **swap.axes**. Argumento de tipo lógico que posibilita el intercambio de ejes. Toma un valor igual a FALSE por defecto, excepto si la matriz de datos es de tipo *CL*.
- **apc.index**. Este argumento permite generar una lista. Si no se realiza ninguna especificación, se generará, automáticamente, una lista de este tipo.
- **sparsity.limits**. Este argumento se corresponde con un vector que posee dos valores ascendentes, y que sirven para reflejar, a su vez, los colores del gráfico. Por defecto, posee asociado el valor $c(1,2)$.
- **cex**. Este argumento posibilita aumentar el tamaño tanto del gráfico como de los símbolos que aparecen en el mismo. Por defecto, toma un valor igual a NULL; en el que, el programa, escoge el tamaño que considera adecuado.

- **pch.** Este argumento representa un vector que está formado por dos valores. Cuando este argumento está compuesto por números enteros reflejan un símbolo; aunque, también puede ser de tipo carácter. Por defecto, posee el valor `c(15,15)`, de tipología cuadrática.
- **main.outer.** Este argumento de la función posibilita que se introduzca un título principal para el gráfico. Este se reproducirá en un margen de la ventana. Por defecto, toma un valor igual a `NULL`.

Los resultados obtenidos mediante la ejecución de esta función se muestran en la figura 4.2.

```
> apc.plot.data.sparsity(base_datos,data.type = "a",main.outer = "Sparsity graph")
```

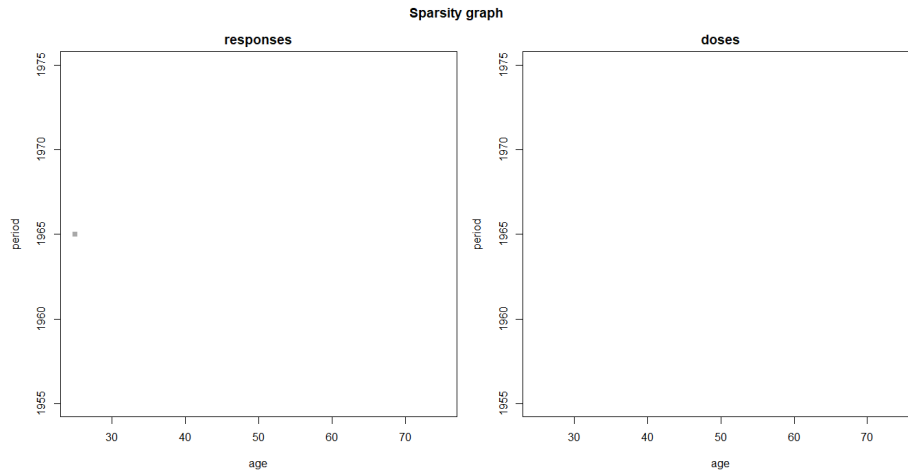


Figura 4.2: Resultados gráficos obtenidos al ejecutar la función `apc.plot.sparsity`.

En este caso, la gráfica de dosis (*doses*) aparece vacía; mientras que, la figura que representa las respuestas (*responses*), únicamente presenta una observación con un valor reducido (en relación con la base de datos).

Por otro lado, la instrucción `apc.plot.data.within` permite representar las series temporales de las matrices edad, periodo y cohorte respecto a uno de los otros dos regresores. En este aspecto, el comando `apc.plot.data.within.all.six` permite generar dichos gráficos en el mismo panel.

Por otra parte, estos dos últimos comandos presentan una notable importancia; ya que, posibilitan la detección de factores temporales no relevantes a la hora de modelar la variable dependiente. En este aspecto, si, al no tener en cuenta uno de los regresores temporales se visualiza un gráfico con líneas paralelas, el factor no considerado podría ser irrelevante; aunque, dicha herramienta debe ser acompañada por métodos analíticos que posibiliten confirmar esta idea.

Como se ha realizado en los casos anteriores, se introducirán los argumentos de esta función:

- **apc.data.list.** Representa la lista de datos con la que se trabajará.
- **data.type.** Argumento opcional de tipo carácter. Puede tomar los siguientes valores: “r” (responses), “d” (dosis), “m” (tasas de mortalidad). Toma, por defecto, el valor “r”.
- **plot.type.** Argumento opcional que posee las siguientes opciones de elección: “awp”, “pwa”, “awc”, “cwa”, “cwp” y “pwc”. En este aspecto, por poner un ejemplo, al seleccionar el caso “awp”, se obtendría la serie con la edad, considerando cada nivel del regresor periodo.
- **average.** Argumento opcional de tipo lógico. Si se selecciona la opción `TRUE`, se realiza el gráfico sobre las medias; mientras que, si se escoge la opción `FALSE`, se obtiene el gráfico en relación con las sumas. En este caso, el valor por defecto es `FALSE`.

- **thin**. Argumento opcional de tipología numérica. Permite agrupar las variables edad, periodo y cohorte en pequeños grupos. Por defecto, se toman grupos en función del tamaño de la matriz de datos.
- **apc.index**. Argumento que permite reflejar una lista; siendo calculado, a su vez, por defecto.
- **ylab**. Argumento que permite modificar la etiqueta asociada al eje Y de las gráficas. Por defecto, toma el valor "".
- **type**. Argumento opcional de tipo carácter que permite modificar los puntos y líneas de los gráficos.
- **log**. Argumento opcional de tipo carácter. Si se escoge la opción "y", el gráfico se representa en escala logarítmica. La opción "" es escogida por defecto.
- **legend**. Valor opcional de tipo lógico que permite representar la leyenda. Por defecto, toma un valor igual a TRUE.
- **lty**. Argumento de carácter vectorial que permite indicar el tamaño que poseerán, gráficamente, los puntos de las columnas de la matriz de datos. Si no se incluye este argumento, se usaría el vector 1:5 (*default option*), de forma cíclica, hasta concluir.
- **col**. Argumento de tipo vectorial que permite indicar los colores que se utilizarán a la hora de realizar el gráfico. Posee una implementación cíclica, al igual que el argumento **lty**.
- **bty**. Argumento de tipo carácter que posibilita escoger el tipo de caja que se dibujará alrededor de la leyenda del gráfico. Posee dos posibles valores de elección: "n" (*default*) y "o".
- **main**. Posibilita la introducción de un título para el gráfico.
- **main.outher**. Valor opcional de tipo carácter. Permite establecer un título principal para el panel con los seis gráficos. Por defecto, toma el valor NULL; generando, de esta manera, un título de forma interna.
- **x**. Este argumento permite indicar la posición en la que se situará la leyenda. Por defecto, toma el valor "topleft".
- **return**. Argumento de tipo lógico que permite obtener, cuando se escoge la opción TRUE, la matriz de datos que ha sido representada.

Los resultados obtenidos para las respuestas (*responses*), dosis (*doses*) y tasas de mortalidad (*rates*); se mostrarán, respectivamente, en las figuras 4.3, 4.4 y 4.5. Estos gráficos han sido obtenidos mediante la implementación de los siguientes comandos:

```
# Para la matriz de respuestas:
> par(mfrow=c(3,2))
> apc.plot.data.within(base_datos,plot.type="awp")
> apc.plot.data.within(base_datos,plot.type="pwa")
> apc.plot.data.within(base_datos,plot.type="awc")
> apc.plot.data.within(base_datos,plot.type="cwa")
> apc.plot.data.within(base_datos,plot.type="cwp")
> apc.plot.data.within(base_datos,plot.type="pwc")

# Para la matriz de dosis:
> par(mfrow=c(3,2))
> apc.plot.data.within(base_datos,plot.type="awp",data.type="d")
> apc.plot.data.within(base_datos,plot.type="pwa",data.type="d")
> apc.plot.data.within(base_datos,plot.type="awc",data.type="d")
> apc.plot.data.within(base_datos,plot.type="cwa",data.type="d")
> apc.plot.data.within(base_datos,plot.type="cwp",data.type="d")
> apc.plot.data.within(base_datos,plot.type="pwc",data.type="d")
```

```

# Para la matriz de ratios de mortalidad:
> par(mfrow=c(3,2))
> apc.plot.data.within(base_datos,plot.type="awp",data.type="m")
> apc.plot.data.within(base_datos,plot.type="pwa",data.type="m")
> apc.plot.data.within(base_datos,plot.type="awc",data.type="m")
> apc.plot.data.within(base_datos,plot.type="cwa",data.type="m")
> apc.plot.data.within(base_datos,plot.type="cwp",data.type="m")
> apc.plot.data.within(base_datos,plot.type="pwc",data.type="m")

```

plots of data using two indices

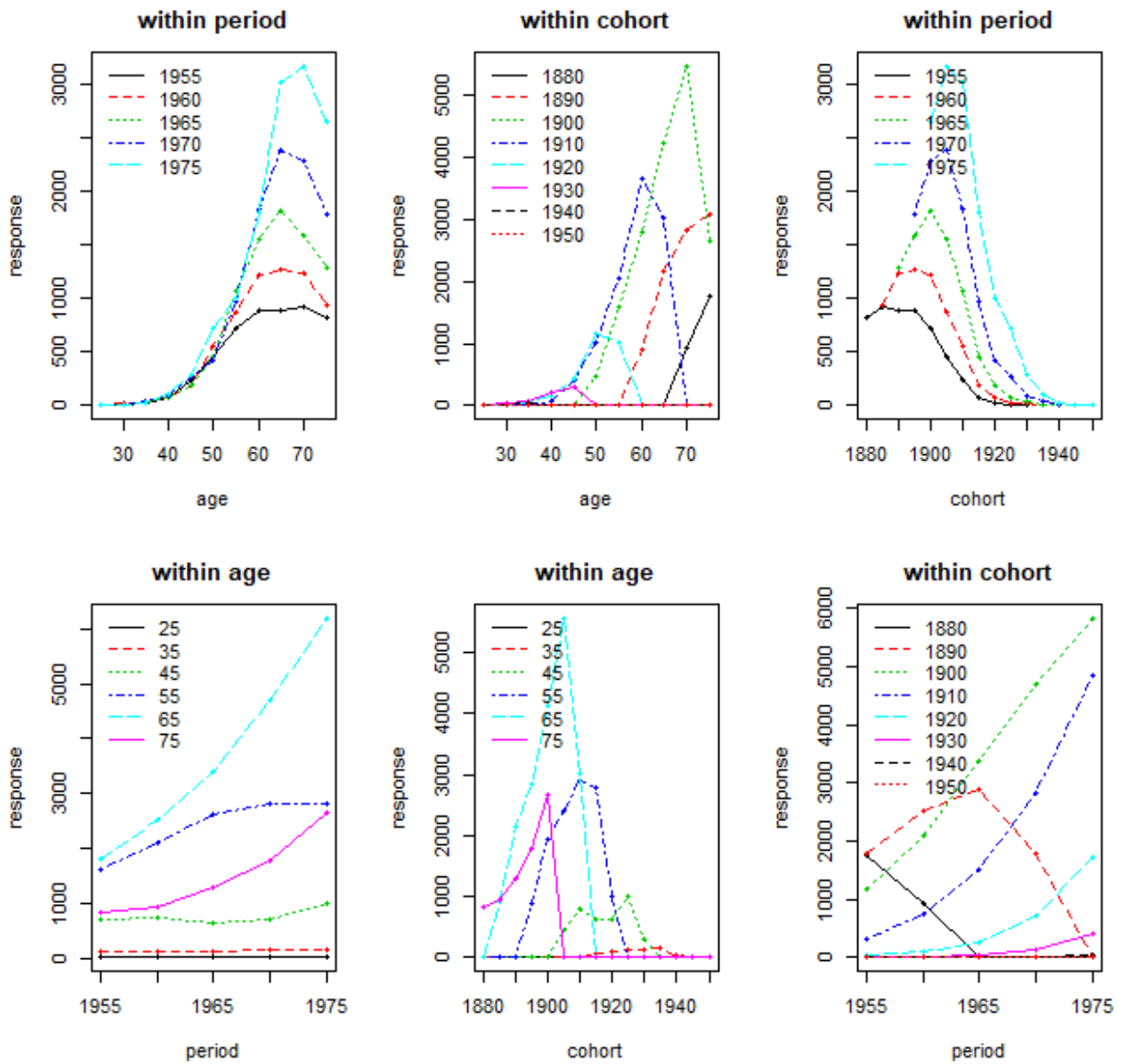


Figura 4.3: Resultados gráficos obtenidos al ejecutar la función *apc.plot.data.within* para las respuestas.

plots of data using two indices

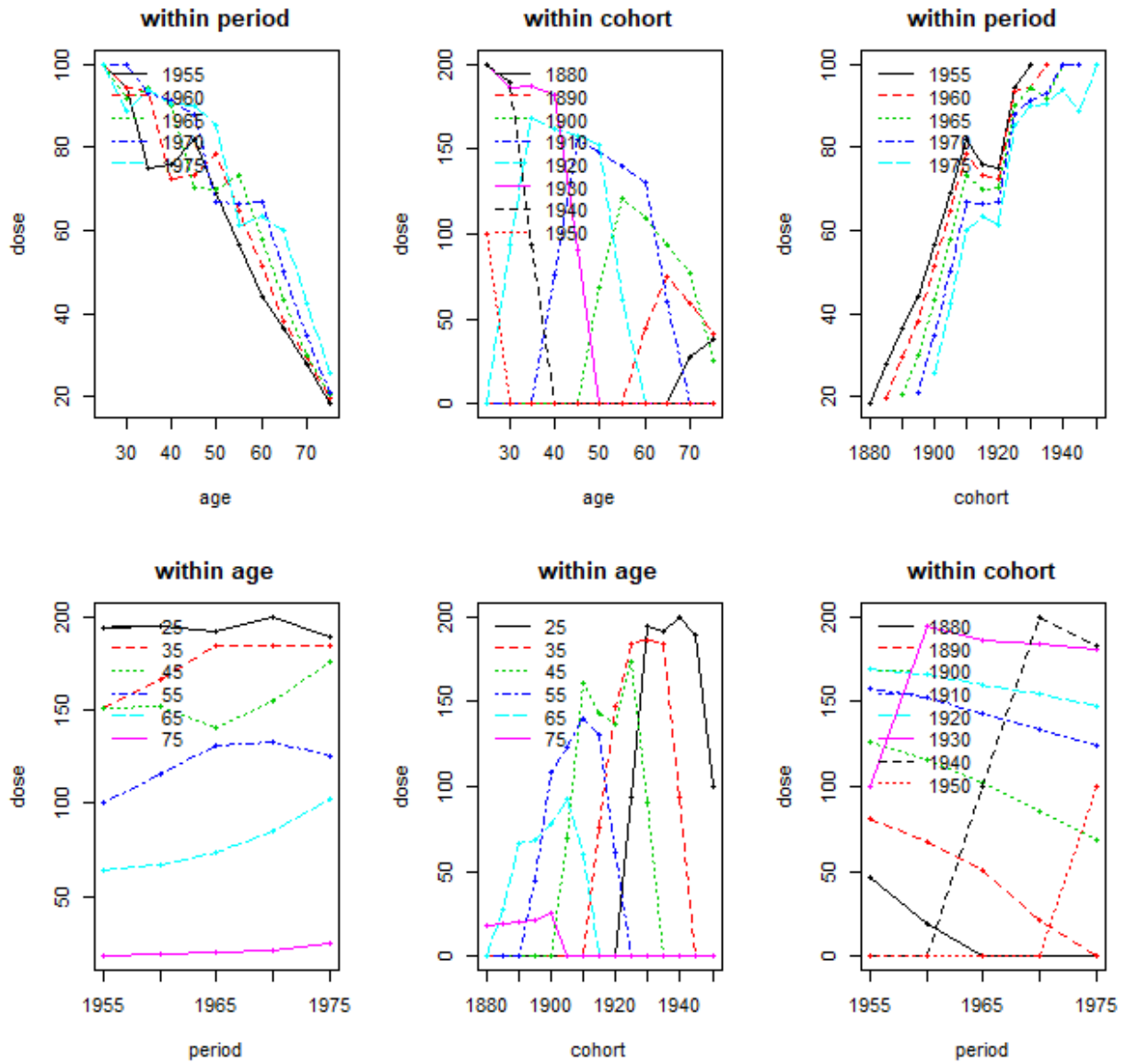


Figura 4.4: Resultados gráficos obtenidos al ejecutar la función *apc.plot.data.within* para las dosis.

plots of data using two indices

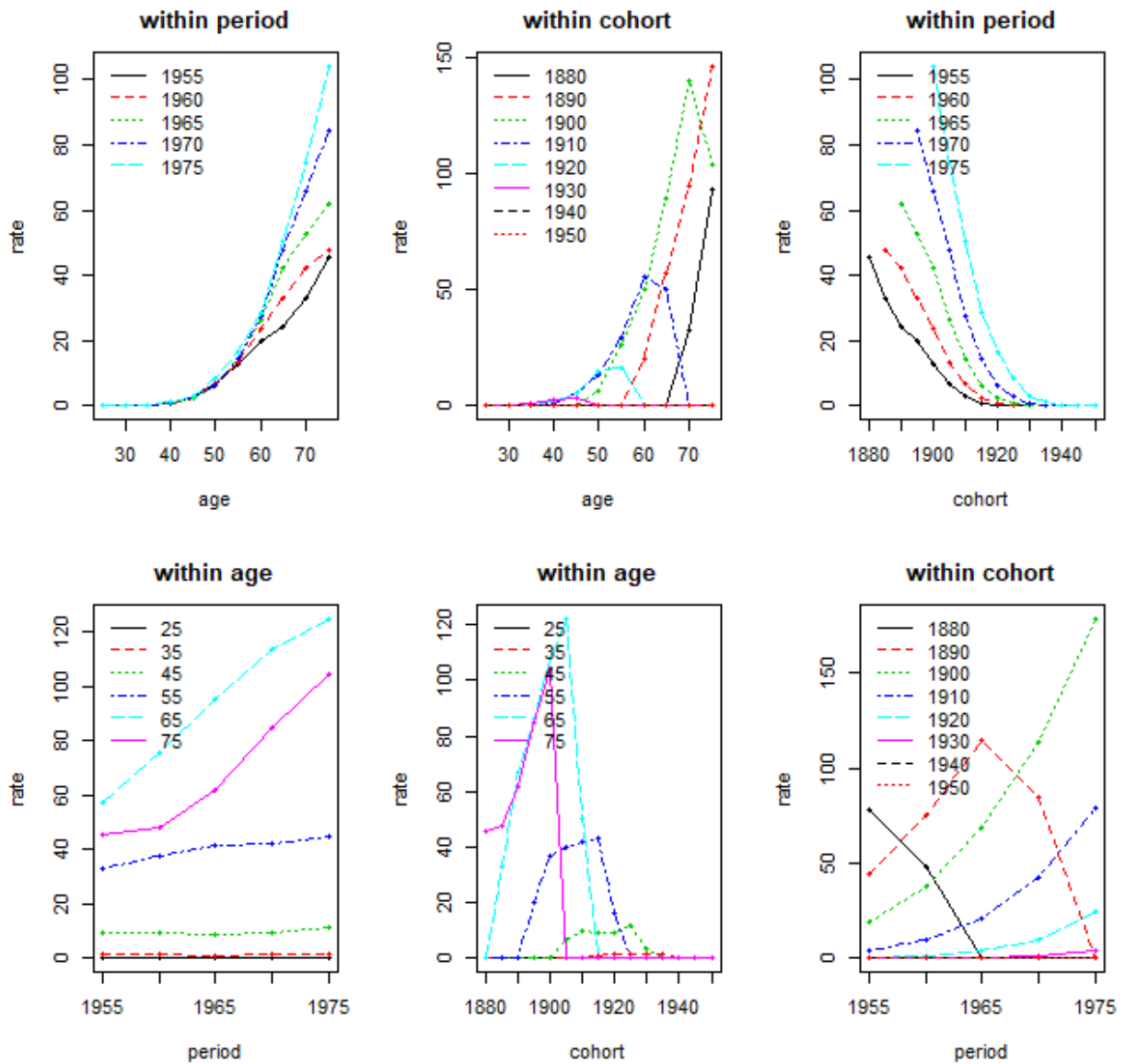


Figura 4.5: Resultados gráficos obtenidos al ejecutar la función `apc.plot.data.within` para los ratios de mortalidad.

Finalmente, se ejecutará la función `apc.plot.data.level`, que muestra la gráfica de nivel de la matriz de datos. Los argumentos de esta función son los siguientes:

- **apc.data.list**. Es el objeto que debe introducirse al comenzar a trabajar con el paquete `apc`.
- **data.type**. Argumento opcional de tipo carácter. Las elecciones disponibles son las siguientes: “r” (*responses*), “d” (*doses*) y “m” (*mortality rates*). La opción “r” es la que se implementa por defecto.
- **rotate**. Argumento opcional de tipo lógico. Si se escoge la opción TRUE, el gráfico gira 90 grados en el sentido de las agujas del reloj (o en el sentido contrario al de las agujas del reloj si se está trabajando con un formato de datos *CL*). Este argumento toma, por defecto, un valor igual a FALSE.
- **apc.index**. Argumento que permite computar una lista que también queda almacenada por defecto.
- **main**. Argumento opcional de tipo carácter que permite dar un título al gráfico.

- **lab**. Este argumento se corresponde con un vector numérico que permite modificar el texto asociado, por defecto, a los ejes del gráfico.
- **contour**. Argumento lógico de carácter opcional. Si toma un valor igual a TRUE, se dibujarán las líneas del contorno del gráfico. Por defecto, el valor de este argumento es FALSE.
- **colorkey**. Permite modificar los colores asociados al gráfico.

Los resultados arrojados por estos gráficos se mostrarán en las figuras 4.6, 4.7 y 4.8, en relación, respectivamente, con las respuestas, dosis y ratios de mortalidad.

```
> apc.plot.data.level(base_datos,data.type = "r",main="Responses")
```

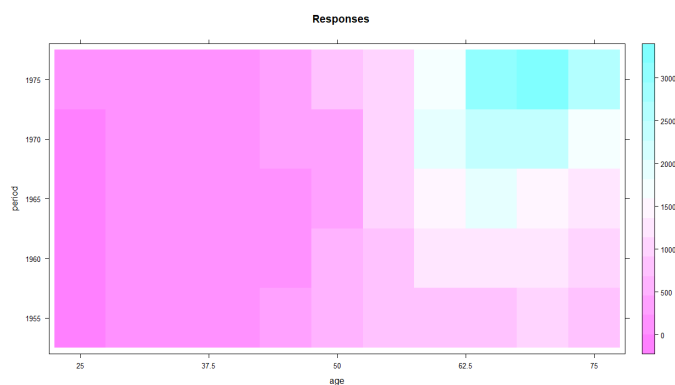


Figura 4.6: Resultados gráficos obtenidos al ejecutar la función *apc.plot.data.level* para las respuestas.

```
> apc.plot.data.level(base_datos,data.type = "d",main="Doses")
```

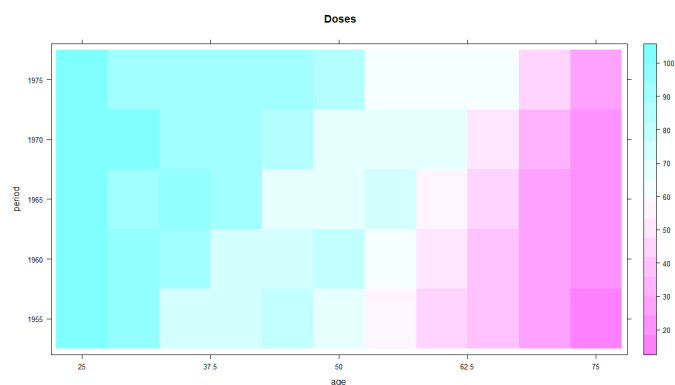


Figura 4.7: Resultados gráficos obtenidos al ejecutar la función *apc.plot.data.level* para las dosis.

```
> apc.plot.data.level(base_datos,data.type = "m",main="Mortality rates")
```

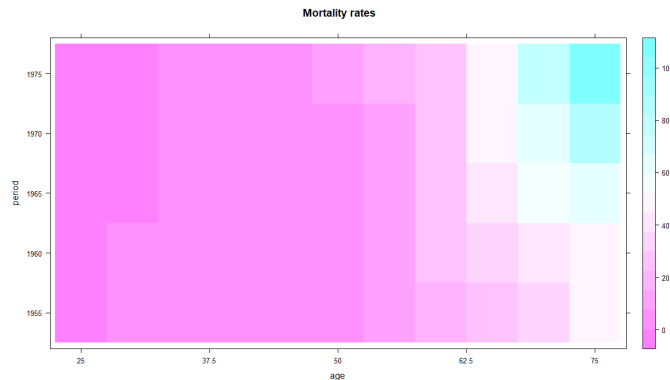


Figura 4.8: Resultados gráficos obtenidos al ejecutar la función `apc.plot.data.level` para las ratios de mortalidad.

4.3. Estimación del modelo

Tras analizar gráficamente los datos, se presentará una estimación de un modelo edad- periodo- cohorte que permita ajustar la matriz de datos con la que se ha ido trabajando a lo largo del ejercicio. Para tal fin, se utilizará la función `apc.fit.model`, que se basa en la parametrización canónica propuesta por Kuang et al. (2008) [5]. Desde un punto de vista teórico, esta metodología fue tratada en el epígrafe 3.3.1.

Por otro lado, la función `apc.fit.model` considera las variables edad, periodo y cohorte; aunque, en caso de que uno de estos regresores no estuviese representado, se trabajaría con las variables que realmente estuviesen presentes; actuando, por tanto, como si se tratase de un submodelo. Además, esta función puede emplearse para analizar las tasas de mortalidad; empleando, para tal fin, un modelo dosis- respuesta (*dose-response model*); aunque, también puede utilizarse para analizar las cifras de mortalidad (sin considerar dosis o exposiciones); utilizando, para ello, un modelo de respuesta pura (*pure response model without doses/exposures*).

Por otra parte, la función `apc.fit.model` utiliza un ajuste a través de la función GLM. De tal forma, las familias correspondientes a las funciones GLM incluyen, por ejemplo, métodos regresivos de Poisson (con sus correspondientes versiones haciendo uso de enlaces logarítmicos- *Poisson regressions with log link*) o regresiones de mínimos cuadrados normales/ gaussianos (*Normal/Gaussian least squares regressions*).

Por otro lado, la función `apc.fit.model` permite obtener una tabla de desviaciones (*deviance table*), que agrupa a los 15 submodelos asociados al modelo general; además de, a las combinaciones referentes al plano lineal. De tal forma, los modelos y submodelos arrojados por esta función son los que se muestran a continuación: APC, AP, AC, PC, Ad, Pd, Cd, A, P, C, t, tA, tP, tC y 1.

Por otra parte, los argumentos de la función `apc.fit.model` son los siguientes:

- **apc.data.list**. Lista que hace referencia al objeto creado al comenzar el análisis de la matriz de datos.
- **model.family**. Argumento de tipo carácter que posee las siguientes opciones de elección (derivadas, a su vez, de la función `glm.fit`):
 - “*poisson.response*”. Con esta opción, se considera que la familia es de tipo Poisson (suponiendo, además, un *link=log*). En este aspecto, únicamente se tendrían en consideración las respuestas (*responses*); utilizando, como método de inferencia, un modelo multinomial.
 - “*od.poisson.response*”. Con esta opción se utiliza, en materia de estimación, una familia de tipo cuasipoisson (haciendo uso, a su vez, de un *link=log*). Sin embargo, a la hora de obtener los errores estándar corregidos, se emplea una familia tipo Poisson (con *link=log*). Por otro lado, en este caso

se consideran, únicamente, las respuestas (responses); utilizando, desde el punto de vista de la inferencia, un modelo de Poisson con sobre dispersión.

- “*poisson.dose.response*”. A través de la selección de esta opción, se considera que la familia es de tipo Poisson (con un *link=log*). Además, las dosis (*doses*) son empleadas como *offset*.
- “*binomial.dose.response*”. Al seleccionar esta opción, se establece que la familia es de tipo Poisson (considerando un *link=log*); devolviendo, a su vez, una regresión logística.
- “*gaussian.rates*”. A través de esta opción, se establece una familia gaussiana; empleando, en este aspecto, un link de tipo *identity*. Además, las variables dependientes serían las tasas de mortalidad (*rates*).
- “*gaussian.response*”. A través de esta opción, se establece una familia gaussiana (*link=identity*). Sin embargo, en este caso se utilizarían, como variables dependientes, las respuestas (*responses*).

- **model.design**. Argumento de tipo carácter que posee las siguientes opciones de elección:
 - *APC*. Modelo edad- periodo- cohorte.
 - *AP*. Modelo edad- periodo. Se encuentra anidado en el modelo APC.
 - *AC*. Modelo edad- cohorte. Se encuentra anidado en el modelo APC.
 - *PC*. Modelo periodo- cohorte. Se encuentra anidado en el modelo APC.
 - *Ad*. Modelo edad-trend que incluye el efecto del regresor edad; además de, dos tendencias de carácter lineal. Se encuentra anidado en los modelos AP y AC.
 - *Pd*. Modelo periodo-trend que incluye el efecto del regresor periodo; además de, dos tendencias de carácter lineal. Se encuentra anidado en los modelos AP y PC.
 - *Cd*. Modelo cohorte-trend que incluye el efecto del regresor cohorte; además de, dos tendencias de carácter lineal. Se encuentra anidado en los modelos AC y PC.
 - *A*. Modelo de tipo edad que se encuentra anidado en el modelo Ad.
 - *P*. Modelo de tipo periodo que se encuentra anidado en el modelo Pd.
 - *C*. Modelo de tipo cohorte que se encuentra anidado en el modelo Cd.
 - *t*. Modelo de tendencia que posee dos tendencias lineales; encontrándose anidado, a su vez, en los modelos Ad, Pd y Cd.
 - *tA*. Modelo que posee una tendencia única relacionada con el índice edad; encontrándose anidado, a su vez, en los modelos A y t.
 - *tP*. Modelo que posee una tendencia única relacionada con el índice periodo; encontrándose anidado en los modelos P y t.
 - *tC*. Modelo que posee una tendencia única relacionada con el índice cohorte; encontrándose anidado, a su vez, en los modelos C y t.
 - 1. Modelo de solo constante. Se encuentra anidado en los modelos tA, tP y tC.
- **model.design.reference**. Argumento opcional de tipo carácter que permite indicar la opción de diseño de referencia asociada a la tabla de desviaciones. En este aspecto, se puede escoger una de las opciones contenidas en el argumento **model.design**; aunque, por defecto, toma un valor de tipo *APC*.
- **apc.index**. Argumento opcional que permite generar una lista que es calculada por defecto. Este argumento es especialmente útil a la hora de implementar simulaciones computacionales.
- **replicate.version.1.3.1**. Argumento opcional de tipo lógico que permite replicar el error asociado a los cálculos de la covarianza.

Una vez explicados los argumentos pertenecientes a la función *apc.fit.model*, esta será aplicada a la base de datos **data.Italian.bladder.cancer**; considerando, en este aspecto, un modelo edad- periodo- cohorte. Los coeficientes canónicos obtenidos se muestran a continuación; siendo, estos últimos, uno de los múltiples elementos obtenidos al ejecutar dicha función:

```

> modelo_apc=apc.fit.model(base_datos,model.family = "poisson.dose.response",
+ model.design = "APC")
> modelo_apc$coefficients.canonical
##              Estimate Std. Error   z value   Pr(>|z|)
## level          1.87648684 0.02828139 66.3505883 0.000000e+00
## age slope       0.76014175 0.03283618 23.1495203 1.470014e-118
## cohort slope    0.03269579 0.02622515  1.2467342 2.124950e-01
## DD_age_35       -0.55065499 0.40664725 -1.3541343 1.756935e-01
## DD_age_40        0.13664203 0.21884545  0.6243768 5.323801e-01
## DD_age_45       -0.05196011 0.13248649 -0.3921918 6.949165e-01
## DD_age_50       -0.10159156 0.08209743 -1.2374511 2.159197e-01
## DD_age_55       -0.16916985 0.05361601 -3.1552112 1.603820e-03
## DD_age_60       -0.13525709 0.03874505 -3.4909512 4.813042e-04
## DD_age_65       -0.08681738 0.03044437 -2.8516723 4.348991e-03
## DD_age_70       -0.09762626 0.02687372 -3.6327777 2.803865e-04
## DD_age_75       -0.02662955 0.02717706 -0.9798538 3.271583e-01
## DD_period_1965  0.01143513 0.03148918  0.3631446 7.164969e-01
## DD_period_1970  0.01900914 0.02772827  0.6855508 4.929964e-01
## DD_period_1975  0.01493432 0.02468704  0.6049457 5.452151e-01
## DD_cohort_1890  0.15804501 0.06198293  2.5498151 1.077801e-02
## DD_cohort_1895  0.03929894 0.04413320  0.8904621 3.732178e-01
## DD_cohort_1900 -0.07836353 0.03405310 -2.3012155 2.137945e-02
## DD_cohort_1905 -0.09985060 0.02798423 -3.5681030 3.595752e-04
## DD_cohort_1910 -0.05478258 0.02801528 -1.9554535 5.052956e-02
## DD_cohort_1915 -0.06493072 0.03267698 -1.9870481 4.691707e-02
## DD_cohort_1920  0.06538341 0.04468068  1.4633487 1.433720e-01
## DD_cohort_1925  0.07563192 0.06125472  1.2347116 2.169379e-01
## DD_cohort_1930 -0.13413655 0.08417885 -1.5934710 1.110546e-01
## DD_cohort_1935  0.12744049 0.13314462  0.9571584 3.384873e-01
## DD_cohort_1940 -0.51208269 0.23735644 -2.1574417 3.097126e-02
## DD_cohort_1945  0.40832896 0.45267592  0.9020337 3.670389e-01
## DD_cohort_1950  1.59638636 0.71890833  2.2205701 2.638009e-02

```

En cuanto a esta última salida, en la columna *estimate* aparecen los coeficientes correspondientes al modelo edad- periodo- cohorte; mientras que, en la columna $\text{Pr}(> |z|)$, se incorpora la significación de dichos coeficientes. De tal forma, los datos contenidos en esta última tabla, muestran las estimaciones correspondientes al nivel y a las pendientes asociadas a los regresores del modelo. En este aspecto, dichas estimaciones permiten definir el plano lineal correspondiente al modelo.

Por otro lado, una vez presentados los coeficientes canónicos del modelo, conviene analizar las transformaciones de probabilidad de respuesta, que fueron calculadas en base a los valores estimados del modelo (estos fueron obtenidos mediante la implementación de la función *apc.fit.model*). Con el objetivo de desarrollar este análisis, se hará uso de la función *apc.plot.fit.pt*, que permite obtener una figura, basada en iconos, que posibilita la visualización de las observaciones en base a las siguientes cuestiones, que se siguen extendiendo en función del código de colores especificado a través de los argumentos de la función:

- Si el icono posee la forma de un círculo negro, la observación se encuentra en los cuartiles centrales (10% – 90% de la distribución).
- Si el icono es un triángulo verde, la observación se encuentra en el 10% de la distribución de las colas. En este aspecto, si el triángulo posee el vértice hacia abajo, la observación se encuentra en la cola izquierda; mientras que, si la punta del triángulo se encuentra situada hacia arriba, la observación se sitúa en la cola de la derecha.
- Si el icono es un triángulo azul, la observación se encuentra dentro del 5% del extremo de las colas. Si el triángulo posee el vértice hacia abajo, la observación se encuentra en la cola izquierda; mientras

que, si la punta del triángulo se encuentra situada hacia arriba, la observación se sitúa en la cola de la derecha.

- Si la figura es un triángulo rojo, la observación se encuentra en los bordes de la distribución de datos.

Por tanto, esta función permite detectar posibles *outliers*, poseyendo los siguientes argumentos:

- **apc.fit.model**. Este argumento hace referencia a la lista de datos que será representada. Es el único argumento obligatorio de la función; habiendo sido obtenido anteriormente.
- **do.plot**. Argumento de tipo lógico que posibilita que no se represente el gráfico. El valor de este argumento, por defecto, es igual a TRUE; por lo que, en este caso, la figura sería representada.
- **do.value**. Argumento opcional de tipo lógico que permite generar un valor al escoger la opción TRUE. Toma un valor, por defecto, igual a FALSE.
- **pch**. Argumento opcional de carácter numérico que permite modificar el tamaño de los iconos del gráfico. Si se consideran los valores por defecto, se tiene la siguiente lista: 21/24/25; representando, respectivamente, el tamaño de los círculos y de los triángulos verdes.
- **col**. Argumento opcional de tipo carácter, o numérico, que permite modificar los colores de los iconos.
- **bg**. Argumento gráfico opcional de tipo carácter o numérico que posee, por defecto, un valor igual a `bg = col`.
- **cex**. Argumento opcional numérico que permite un aumento del tamaño de la figura. Dicho argumento es computado, de manera interna, por defecto.
- **main**. Valor opcional de tipo carácter que permite introducir un título al gráfico. Este argumento es computado, internamente, por defecto.

Tras introducir los argumentos de la función *apc.plot.fit.pt*, se procederá con la implementación de la misma:

```
> apc.plot.fit.pt(modelo_apc)
```

Los resultados obtenidos se muestran en la figura 4.9.

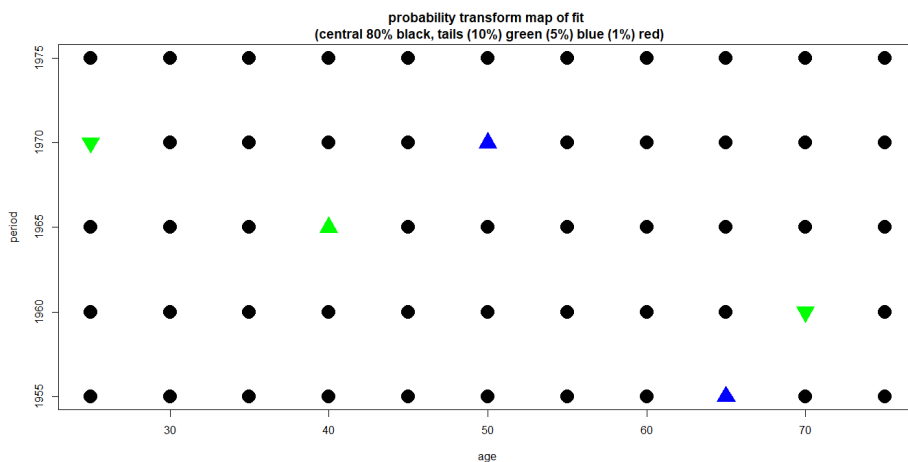


Figura 4.9: Resultados gráficos obtenidos al ejecutar la función *apc.plot.fit.pt*.

Por otro lado, tras crear esta última figura, conviene obtener las gráficas de las estimaciones asociadas a los parámetros del modelo. Para ello, se puede utilizar la función *apc.plot.fit*, que posee los siguientes argumentos:

- **apc.fit.model**. Hace referencia a la lista de datos obtenida a la hora de estimar el modelo especificado. Este argumento fue definido anteriormente.
- **scale**. Argumento opcional de tipo lógico que toma el valor FALSE por defecto; permitiendo, de esta manera, que se represente la escala de la función de enlace. Si se selecciona la opción TRUE, se usaría la escala inversa de la función de enlace.
- **sdv.at.zero**. Argumento opcional de tipo lógico que permite representar las desviaciones estándar, alrededor de las estimaciones, si se selecciona la opción FALSE. Si se escoge la opción TRUE, que viene por defecto, las desviaciones estándar se representan alrededor de cero.
- **type**. Argumento opcional de tipo carácter que posee las siguientes opciones de elección:
 - *detrend*. Al seleccionar esta opción, cada serie comienza y termina en cero; ya que, la última fila correspondiente a las representaciones gráficas, se corresponde con las dobles sumas de las dobles diferencias. De tal forma, en la fila central de las figuras aparecerá el nivel pertinente y, como máximo, dos tendencias lineales.
 - *sum.sum*. Al escoger esta opción, la última fila de los gráficos se corresponde con las dobles sumas de las dobles diferencias ampliadas; siguiendo, pues, el proceso de estimación de Nielsen (2015) [129]. Por otro lado, el nivel pertinente y, como máximo, dos tendencias lineales, se sitúan en la fila central del gráfico.
- **include.linear.plane**. Argumento opcional de tipo lógico que permite incluir, al escoger la opción TRUE, la representación del plano lineal asociado al modelo. Toma un valor, por defecto, igual a TRUE.
- **include.double.differences**. Argumento opcional de tipo lógico que permite incluir, al escoger la opción TRUE, la representación de las dobles sumas de las dobles diferencias asociadas al modelo. En este aspecto, la opción TRUE es la ejecutada por defecto.
- **sub.plot**. Este argumento opcional de tipo carácter permite escoger los gráficos que serán mostrados. El valor de este argumento puede tomar una letra entre la *a* y la *i*, en función del gráfico que se desea obtener. Por defecto toma un valor igual a NULL; por lo que, aparecerán todas las figuras.
- **main.outer**. Argumento opcional de tipo carácter que permite generar un título principal para el gráfico.
- **main.sub**. Argumento opcional de tipo vectorial que permite establecer títulos para los distintos subgráficos. Dicho vector posee nueve caracteres.
- **cex**. Argumento opcional que permite modificar el tamaño del texto. Toma un valor igual a 1 por defecto.
- **cex.axis**. Argumento opcional que permite controlar el tamaño de las anotaciones que aparecen en los ejes de los gráficos. Toma un valor igual a 1 por defecto.
- **cex.lab**. Argumento opcional de tipo gráfico que permite aumentar las etiquetas de los ejes de los gráficos.
- **cex.main**. Argumento opcional de tipo gráfico que permite aumentar el título principal del gráfico.
- **cex.main.outer**. Argumento opcional que permite ampliar el título principal externo que aparece al mostrar la matriz de gráficos. Por defecto, toma un valor igual a 1.2.
- **line.main**. Argumento opcional que permite modificar la posición de los títulos principales pertenecientes a los gráficos individuales. Toma un valor, por defecto, de 0.5.
- **line.main.outer**. Argumento opcional que permite especificar la posición de la línea del título principal externo, si se decide que este sea mostrado. Toma un valor, por defecto, igual a NULL.
- **las**. Argumento opcional numérico que permite modificar el estilo de las etiquetas de los ejes.

- **mar.** Argumento opcional que permite definir la cifra de líneas de margen a especificar en los cuatro lados de la parcela de gráficos. De forma predeterminada, toma los valores $c(4,3,2,0)$ para una matriz de parcelas; mientras que, las cifras $c(4,4,3,1)$ son aplicadas, predeterminadamente, a una única gráfica.
- **oma.** Argumento opcional que permite modificar el tamaño de los márgenes exteriores de los gráficos correspondientes a las líneas de texto. De esta manera, de forma predeterminada, toma un valor de $c(0,0,5,1)$ para una matriz de parcelas; mientras que, toma un valor de $c(0,0,0,0)$, para una única parcela.
- **mgp.** Argumento opcional que permite modificar el parámetro de trazado de la línea de margen para el título, etiqueta y línea del eje. El valor predeterminado es $c(2,1,0)$.
- **vec.xlab.** Argumento opcional que permite controlar los títulos, para el eje X, de los gráficos.

Si se aplica esta función a los datos, se obtienen los resultados mostrados en las figuras 4.10 y 4.11. De tal forma, en las subgráficas (a), (b) y (c) se introducen las estimaciones de los parámetros asociados a las segundas diferencias ($\sum^2 \alpha_i, \sum^2 \beta_j$ y $\sum^2 \gamma_k$). Por otro lado, en relación con la figura 4.10, al escoger la opción *detrend*, se eliminaría el ruido asociado a las estimaciones.

```
> apc.plot.fit(modelo_apc, type = "detrend")
```

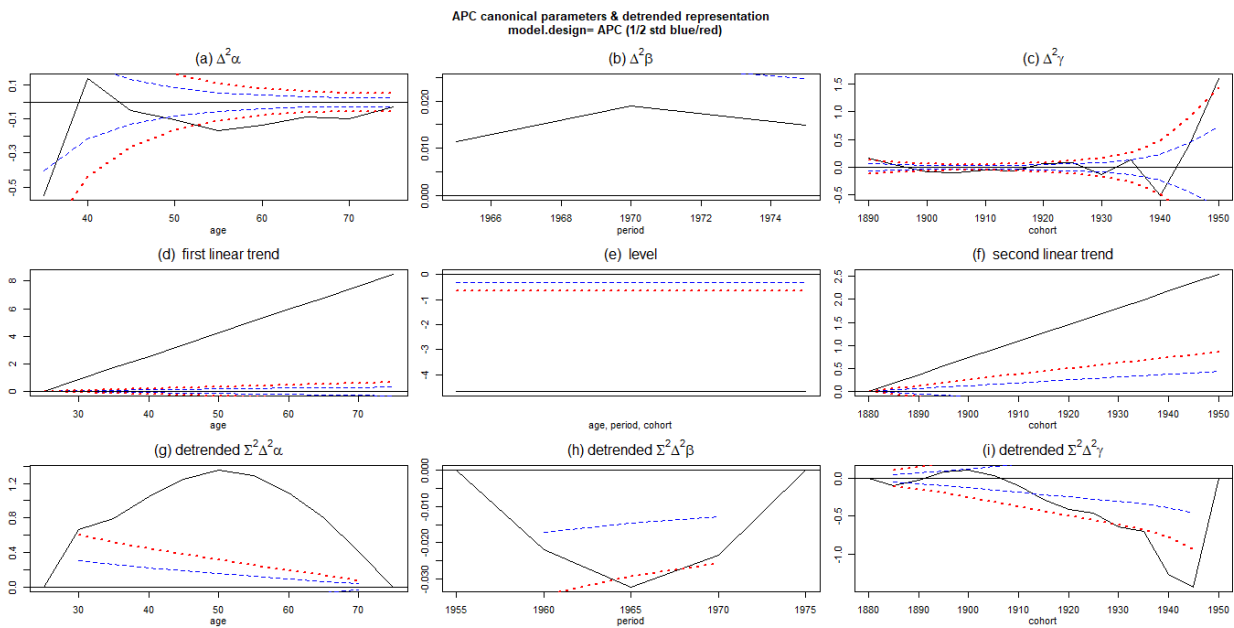


Figura 4.10: Resultados gráficos obtenidos al ejecutar la función *apc.plot.fit* escogiendo la opción *detrend*.


```
> apc.plot.fit(modelo_apc, type = "sum.sum")
```

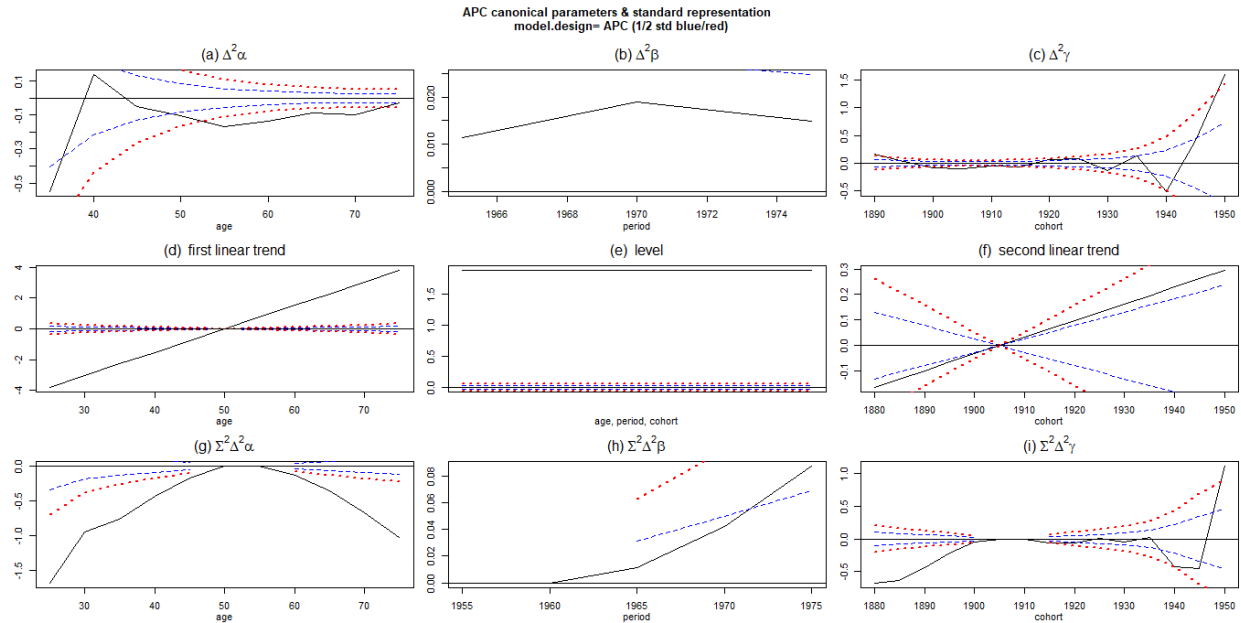


Figura 4.11: Resultados gráficos obtenidos al ejecutar la función *apc.plot.fit* escogiendo la opción *sum.sum*.

Por otro lado, la función *apc.identify* permite obtener los efectos temporales que han sido identificados ad hoc. De tal forma, esta función posee, únicamente, el argumento **apc.fit.model**, que fue descrito anteriormente.

Por otra parte, en cuanto a las salidas que se obtienen al implementar la función *apc.identify*, destacan los *coefficients.ssdd*, que se corresponden con los datos de la matriz de las dobles sumas, de las diferencias segundas, pertenecientes al parámetro canónico normalizado. De tal forma, se obtendrá, para cada parámetro: el valor de su coeficiente, la desviación estándar asociada a este, el valor *z* y el *p* valor. Por tanto, si se aplica esta función al modelo ejecutado, se obtienen los siguientes datos:

```
> apc.identify(modelo_apc)$coefficients.ssdd
```

##	Estimate	Std. Error	z value	Pr(> z)
## level	1.87648684	0.02828139	66.3505883	0.000000e+00
## age slope	0.76014175	0.03283618	23.1495203	1.470014e-118
## cohort slope	0.03269579	0.02622515	1.2467342	2.124950e-01
## SS_DD_age_25	-1.68546674	0.34799016	-4.8434322	1.276153e-06
## SS_DD_age_30	-0.94873226	0.19005118	-4.9919829	5.976254e-07
## SS_DD_age_35	-0.76265277	0.13067077	-5.8364451	5.332634e-09
## SS_DD_age_40	-0.43993125	0.08720118	-5.0450146	4.534867e-07
## SS_DD_age_45	-0.16916985	0.05361601	-3.1552112	1.603820e-03
## SS_DD_age_50	0.00000000	NA	NA	NA
## SS_DD_age_55	0.00000000	NA	NA	NA
## SS_DD_age_60	-0.13525709	0.03874505	-3.4909512	4.813042e-04
## SS_DD_age_65	-0.35733156	0.06317220	-5.6564687	1.545192e-08
## SS_DD_age_70	-0.67703229	0.08884289	-7.6205573	2.525829e-14
## SS_DD_age_75	-1.02336257	0.11495337	-8.9024150	5.464453e-19
## SS_DD_period_1955	0.00000000	NA	NA	NA
## SS_DD_period_1960	0.00000000	NA	NA	NA
## SS_DD_period_1965	0.01143513	0.03148918	0.3631446	7.164969e-01

```

## SS_DD_period_1970  0.04187939 0.04966090  0.8433072  3.990567e-01
## SS_DD_period_1975  0.08725797 0.06887270  1.2669458  2.051747e-01
## SS_DD_cohort_1880 -0.67176301 0.10146742 -6.6204799  3.580346e-11
## SS_DD_cohort_1885 -0.63611025 0.07921295 -8.0303825  9.716925e-16
## SS_DD_cohort_1890 -0.44241248 0.06145619 -7.1988272  6.073265e-13
## SS_DD_cohort_1895 -0.20941576 0.04438280 -4.7183995  2.377073e-06
## SS_DD_cohort_1900 -0.05478258 0.02801528 -1.9554535  5.052956e-02
## SS_DD_cohort_1905  0.00000000          NA          NA          NA
## SS_DD_cohort_1910  0.00000000          NA          NA          NA
## SS_DD_cohort_1915 -0.06493072 0.03267698 -1.9870481  4.691707e-02
## SS_DD_cohort_1920 -0.06447803 0.05122096 -1.2588212  2.080949e-01
## SS_DD_cohort_1925  0.01160657 0.07005291  0.1656829  8.684065e-01
## SS_DD_cohort_1930 -0.04644538 0.09614248 -0.4830890  6.290325e-01
## SS_DD_cohort_1935  0.02294317 0.13344826  0.1719256  8.634960e-01
## SS_DD_cohort_1940 -0.41975096 0.21356710 -1.9654290  4.936461e-02
## SS_DD_cohort_1945 -0.45411614 0.34745735 -1.3069694  1.912231e-01
## SS_DD_cohort_1950  1.10790504 0.45271168  2.4472641  1.439453e-02

```

Por otro lado, también se pueden obtener las estimaciones asociadas a los coeficientes de la parametrización realizada a través del método *detrend*. Estas estimaciones generan una matriz con los coeficientes de las dobles sumas, de las segundas diferencias, obtenidas a partir de la normalización del primer y del último valor estimado. Por tanto, para obtener dicha matriz, se deben implementar las siguientes instrucciones:

```
> apc.identify(modelo_apc)$coefficients.ssdd
```

```

##           Estimate Std. Error    z value    Pr(>|z|)
## level      -4.66307555 0.32033574 -14.5568383 5.284596e-48
## age slope   0.84816666 0.03184519  26.6340584 2.738348e-156
## cohort slope 0.18162943 0.03080123   5.8968235 3.705659e-09
## SS_DD_age_25 0.00000000          NA          NA          NA
## SS_DD_age_30 0.67052406 0.30079817   2.2291494 2.580397e-02
## SS_DD_age_35 0.79039314 0.26009416   3.0388731 2.374649e-03
## SS_DD_age_40 1.04690424 0.22378111   4.6782512 2.893319e-06
## SS_DD_age_45 1.25145523 0.19137282   6.5393572 6.178382e-11
## SS_DD_age_50 1.35441466 0.15963875   8.4842475 2.171192e-17
## SS_DD_age_55 1.28820424 0.12791829  10.0705239 7.457975e-24
## SS_DD_age_60 1.08673673 0.09638382  11.2750946 1.741913e-29
## SS_DD_age_65 0.79845184 0.06526031  12.2348760 2.024136e-34
## SS_DD_age_70 0.41254070 0.03543313  11.6427949 2.496966e-31
## SS_DD_age_75 0.00000000          NA          NA          NA
## SS_DD_period_1955 0.00000000          NA          NA          NA
## SS_DD_period_1960 -0.02181449 0.01721817  -1.2669458 2.051747e-01
## SS_DD_period_1965 -0.03219386 0.01461343  -2.2030324 2.759246e-02
## SS_DD_period_1970 -0.02356409 0.01280427  -1.8403309 6.571967e-02
## SS_DD_period_1975 0.00000000          NA          NA          NA
## SS_DD_cohort_1880 0.00000000          NA          NA          NA
## SS_DD_cohort_1885 -0.09146638 0.05165645  -1.7706672 7.661607e-02
## SS_DD_cohort_1890 -0.02488776 0.07161387  -0.3475271 7.281954e-01
## SS_DD_cohort_1895  0.08098981 0.09781464   0.8279927 4.076746e-01
## SS_DD_cohort_1900  0.10850385 0.12603895   0.8608755 3.893066e-01
## SS_DD_cohort_1905  0.03616728 0.15574748   0.2322174 8.163692e-01
## SS_DD_cohort_1910 -0.09095187 0.18562652  -0.4899724 6.241534e-01
## SS_DD_cohort_1915 -0.28300174 0.21591392  -1.3107156 1.899539e-01
## SS_DD_cohort_1920 -0.40966820 0.24651093  -1.6618663 9.653959e-02
## SS_DD_cohort_1925 -0.46070274 0.27690228  -1.6637737 9.615766e-02
## SS_DD_cohort_1930 -0.64587383 0.30723031  -2.1022465 3.553169e-02

```

```
## SS_DD_cohort_1935 -0.70360443 0.33934944 -2.0733921 3.813581e-02
## SS_DD_cohort_1940 -1.27341771 0.38834684 -3.2790732 1.041486e-03
## SS_DD_cohort_1945 -1.43490204 0.46391829 -3.0930059 1.981402e-03
## SS_DD_cohort_1950 0.00000000 NA NA NA
```

4.4. Inferencia sobre el modelo

Tras estimar el modelo edad- periodo- cohorte es conveniente verificar si, este es el modelo que permite ajustar, en mayor medida, el comportamiento de la base de datos con la que se está trabajando. Por tanto, el objetivo consiste en obtener el modelo que permita ajustar, de la manera más adecuada, la base de datos disponible. Para tal fin, se generará un cuadro de desviaciones (*deviance table*) que permita contrastar la adecuación de los distintos modelos y submodelos que puedan ser implementados. En este aspecto, dicha tabla puede obtenerse a través de la función *apc.fit.table*, que posee los mismos argumentos que la función *apc.fit.model*, y que presenta, a su vez, las siguientes salidas:

- **-2logL**. Esta salida se corresponde con la *deviance* para el modelo de Poisson o el modelo Binomial. Desde un punto de vista estadístico, se corresponde con la diferencia de -2 veces el logaritmo de la función de verosimilitud entre el modelo estimado (que se encuentra, a su vez, normalizado a cero) y el modelo saturado.
- **df.residual**. Se corresponde con los grados de libertad para los residuos: $nrow \times (ncol - ndim)$. Si la familia de modelos escogida es de tipo *Poisson.response*, los grados de libertad deben tener un grado menor al estipulado a partir de la fórmula introducida anteriormente.
- **prob(>chi_sq)**. Salida que se corresponde con el *p-value* referente a las deviancias ($-2\log(L)$). Esta cifra no aparecerá en el caso Gaussiano; ya que, no se podrá obtener el modelo saturado.
- **LR vs APC**. Esta salida enfrenta al estadístico de la ratio de verosimilitud frente al modelo edad- periodo- cohorte.
- **df**. Grados de libertad obtenidos en contra del modelo edad- periodo- cohorte.
- **prob(>chi_sq)**. Esta salida se corresponde con el *p-value* del estadístico log- razón de verosimilitud.
- **aic**. Esta salida se corresponde con el criterio de información de Akaike (*Akaike information criteria*); por lo que, se obtiene a partir de la función GLM; adoptando, a su vez, diferentes valores en función de la familia de modelos considerada. De tal forma, en el caso de las familias *poisson.dose.response* y *binomial.dose.response*, la dispersión debe ser fijada a uno; mientras que, el número de parámetros se corresponde con la cifra de coeficientes. Por otro lado, los modelos *poisson.response* deben ser condicionados en función del nivel y de la cifra de parámetros (número al que deben ser restadas dos unidades), con el objetivo de obtener el AIC más adecuado. Por otra parte, la dispersión debe ser estimada en función de la *deviance* residual, en el caso de las familias *gaussian.response* y *gaussian.dose.response*.
- **F**. Únicamente puede obtenerse al seleccionar la familia *od.poisson.response*. En este aspecto, se corresponde con la ratio de desviaciones (*deviance*) para el submodelo, dividido por los grados de libertad de las desviaciones del modelo general; y dividido, a su vez, por los grados de libertad correspondientes.
- **prob(>F)**. Solo puede obtenerse cuando se selecciona la familia *od.poisson.response*. Se corresponde con los grados de libertad obtenidos a partir de las diferencias existentes entre el submodelo y el modelo general y entre este último modelo y el modelo saturado.

Cuando se aplica esta función a la base de datos con la que se ha estado trabajando a lo largo de este epígrafe, se obtienen los siguientes resultados:

```
> deviance_table=apc.fit.table(base_datos,"poisson.dose.response")
> deviance_table
```

##	deviance	df.residual	prob(>chi_sq)	LR vs.APC	df vs.APC	prob(>chi_sq)
## APC	33.179	27	0.191	NaN	NaN	NaN
## AP	512.514	40	0.000	479.335	13	0.000
## AC	39.390	30	0.117	6.211	3	0.102
## PC	1146.649	36	0.000	1113.470	9	0.000
## Ad	518.543	43	0.000	485.364	16	0.000
## Pd	4041.373	49	0.000	4008.194	22	0.000
## Cd	1155.629	39	0.000	1122.450	12	0.000
## A	2223.800	44	0.000	2190.621	17	0.000
## P	84323.944	50	0.000	84290.765	23	0.000
## C	23794.205	40	0.000	23761.026	13	0.000
## t	4052.906	52	0.000	4019.727	25	0.000
## tA	5825.158	53	0.000	5791.979	26	0.000
## tP	84325.758	53	0.000	84292.579	26	0.000
## tC	33446.796	53	0.000	33413.617	26	0.000
## 1	87313.678	54	0.000	87280.499	27	0.000
##	aic					
## APC	487.624					
## AP	940.958					
## AC	487.835					
## PC	1583.094					
## Ad	940.988					
## Pd	4451.818					
## Cd	1586.074					
## A	2644.245					
## P	84732.389					
## C	24222.650					
## t	4457.351					
## tA	6227.602					
## tP	84728.203					
## tC	33849.241					
## 1	87714.123					

Teniendo en cuenta estas salidas, el modelo que presenta una mejor adecuación a la base de datos es el modelo edad- periodo- cohorte (fila *APC*).

Capítulo 5

Aplicación de un *APC model* a las cifras y las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019)

En este epígrafe, se aplicará un modelo edad- periodo- cohorte a las cifras y las tasas de mortalidad por suicidio, por 100000 habitantes, de Inglaterra y Gales correspondientes al periodo 1985- 2019. En este aspecto, en primer lugar, se analizarán las cifras brutas de suicidios; estudiándose, posteriormente, las tasas de mortalidad por suicidio.

5.1. Aplicación de un *APC model* a las cifras de suicidios totales

En primer lugar, se aplicará un modelo edad- periodo- cohorte con el objetivo de estudiar las cifras brutas de suicidios. Para ello, se utilizará el paquete *apc*, cuyo funcionamiento ha sido descrito en el capítulo 4. En este aspecto, se irán describiendo los resultados obtenidos, con el objetivo de estudiar este problema de salud pública.

Por otro lado, como se ha señalado anteriormente, las bases de datos a utilizar poseen una estructura edad- periodo (age- period array). En este aspecto, se utilizarán las frecuencias de fallecidos por suicidio (*responses*) junto con el grupo de exposición asociado a dicha enfermedad; siendo, en este caso, dicho grupo, el total poblacional (*doses*). De tal forma, debido a la naturaleza de los datos a analizar, se utilizará un modelo *poisson- dose- response*.

Una vez señaladas estas cuestiones preliminares, se comenzará con el análisis de las cifras brutas de suicidios.

5.1.1. Creación del objeto de datos

En primer lugar, se deben introducir los datos que se utilizarán en el estudio en el software R. Para ello, se ha utilizado la función *read_excel*; que se encuentra disponible, a su vez, en la librería *readxl*; ya que, los datos del estudio se encontraban contenidos en un archivo *.xls*. Una vez introducidos los datos, estos han sido depurados y transformados en matrices; siguiendo, para ello, los pasos contenidos en el anexo de este documento (apéndice A.1.1).

Una vez se dispone de las bases de datos a analizar, se debe crear el correspondiente objeto de tipología *apc.data.list*. Debido a la naturaleza de los datos; se han seguido, previamente, los pasos contenidos en Fannon y Nielsen (2020) (pp. 9-10) [8], con el objetivo de obtener la matriz *doses*. De tal forma, dichos pasos se muestran a continuación, en relación con la base de datos de suicidios totales:

```
> library(apc) # Se carga la librería
# Se transforman los dataframes en matrices (para trabajar con el paquete apc):
> numero_suicidios_totales_matrix=as.matrix(t_Numero_Suicidios_Totales)
> Tasas_Suicidios_Totales_matrix=as.matrix(t_Tasas_Suicidios_Totales)
> numero_suicidios_totales=apc.data.list(response = numero_suicidios_totales_matrix,
dose =numero_suicidios_totales_matrix/Tasas_Suicidios_Totales_matrix ,data.format = "AP",
age1=15, per1 = 1985, unit = 5, label = "Número de suicidios totales")
> numero_suicidios_totales
## $response
##      1985-1989 1990-1994 1995-1999 2000-2004 2005-2009 2010-2014 2015-2019
## 15-19      1009      847      852      803      656      667      901
## 20-24      2362      2606      2096      1701      1467      1610      1777
## 25-29      2290      2909      2883      2236      1787      1767      2093
## 30-34      2328      2655      3138      2778      2100      1989      2286
## 35-39      2438      2471      2656      2962      2706      2347      2238
## 40-44      2425      2603      2472      2572      2782      2993      2557
## 45-49      2196      2432      2353      2294      2442      3044      3085
## 50-54      2138      1915      1979      2122      2081      2712      2910
## 55-59      2157      1641      1529      1779      1883      2033      2279
## 60-64      2157      1454      1189      1142      1379      1561      1640
## 65-69      1961      1420      1065      985      923      1080      1234
## 70-74      1703      1214      992      904      733      839      988
## 75-79      1525      1033      894      829      651      680      735
## 80-84      1026      856      679      620      629      604      617
## 85-89      411      474      449      425      371      415      441
##
## $dose
##      1985-1989 1990-1994 1995-1999 2000-2004 2005-2009 2010-2014 2015-2019
## 15-19 188.24627 157.43494 154.90909 164.54918 175.40107 173.69792 165.01832
## 20-24 206.46853 190.21898 159.75610 161.38520 179.77941 189.85849 186.65966
## 25-29 191.15192 206.01983 192.71390 171.20980 181.23732 192.90393 200.47893
## 30-34 169.67930 189.91416 206.17608 198.71245 180.10292 188.70968 198.09359
## 35-39 176.41100 169.47874 189.17379 207.13287 201.63934 182.22050 190.30612
## 40-44 167.70401 176.11637 169.54733 189.67552 207.61194 201.68464 182.90415
## 45-49 140.58899 165.21739 174.03846 168.18182 187.84615 206.23306 201.89791
## 50-54 133.95990 137.96830 163.82450 171.68285 165.68471 187.03448 205.50847
## 55-59 133.80893 130.03170 134.83245 159.40860 166.93262 162.90064 184.08724
## 60-64 134.30884 126.43478 124.63312 129.77273 152.88248 161.26033 159.53307
## 65-69 121.95274 122.20310 116.52079 116.70616 120.81152 146.34146 155.02513
## 70-74 101.61098 105.38194 106.20985 103.66972 105.01433 111.27321 136.08815
## 75-79 82.43243 79.95356 85.14286 87.07983 87.50000 91.15282 98.26203
## 80-84 52.13415 56.83931 56.20861 62.24900 64.97934 68.32579 73.27791
## 85-89 24.23349 28.97311 32.53623 32.89474 37.70325 41.83468 45.27721
##
## $data.format
## [1] "AP"
##
## $age1
## [1] 15
##
## $per1
```

```
## [1] 1985
##
## $coh1
## [1] 5
##
## $unit
## [1] 5
##
## $per.zero
## NULL
##
## $per.max
## NULL
##
## $time.adjust
## [1] 0
##
## $label
## [1] "Numero de suicidios totales"
##
## $n.decimal
## NULL
```

5.1.2. Exploración de los datos

Una vez introducido el objeto en formato *apc.data.list*, se procederá con el análisis gráfico correspondiente a las cifras de suicidios, de Inglaterra y Gales, pertenecientes al periodo 1985- 2019. Para realizar dicho análisis, se introducirá el siguiente comando:

```
> apc.plot.data.all(numero_suicidios_totales) # El objeto numero_suicidio_totales fue
construido anteriormente
```

Los resultados gráficos obtenidos a través de esta instrucción se irán mostrando, progresivamente, junto a una breve descripción de los mismos. De tal forma, en primer lugar, en la figura 5.1 se muestran 9 imágenes que representan las sumas de los datos en función de los índices edad, periodo y cohorte.

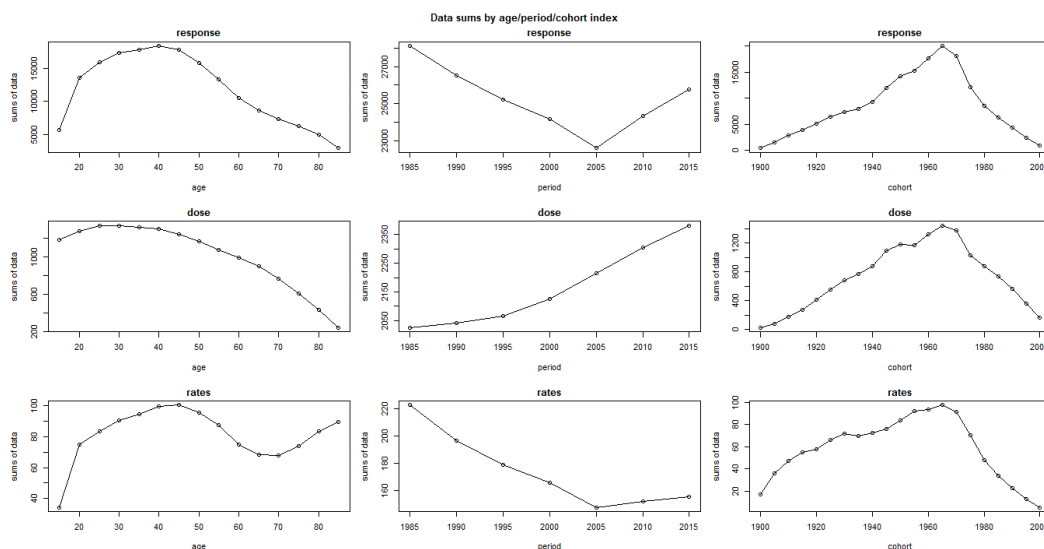


Figura 5.1: Representación de la suma de los datos en función de los índices edad, periodo y cohorte.

De tal forma, la primera fila de gráficos contenida en la figura 5.1, representa la evolución de las muertes por suicidio (*responses*) en función de las variables independientes del modelo. En este aspecto, respecto al regresor edad, se observa un aumento del número de suicidios entre los 15 y los 44 años; reduciéndose esta cifra, de forma significativa, a partir de dicha edad. Por otro lado, en el caso del regresor periodo, se observa una disminución del número de suicidios entre 1985 y 2005, año a partir del cual se produce un aumento de la cifra de suicidios que continúa hasta el final del periodo de análisis. Por otra parte, respecto a la variable cohorte, se observa un crecimiento de los fallecimientos por suicidio entre las cohortes nacidas entre 1900 y 1965; mostrando un comportamiento decreciente, dicha serie, a raíz de esta cohorte de nacimiento.

Por otro lado, en las imágenes que aparecen en la segunda fila de la figura 5.1, se presenta la evolución de las dosis (*doses*); representando, en este caso, a la población objeto de estudio. De tal forma, en relación con la variable edad, la población alcanza el máximo en torno a los 25- 30 años; poseyendo, a partir de este punto, una tendencia decreciente. Por otro lado, respecto al regresor periodo, la población presenta una tendencia creciente, ilustrando el incremento del número de habitantes, que han experimentado Inglaterra y Gales, a lo largo del periodo de análisis. Finalmente, en relación con la variable cohorte, se observa una tendencia creciente entre 1900 y 1965.

Por otra parte, en las imágenes pertenecientes a la tercera fila de la figura 5.1, se presenta la evolución de las tasas de mortalidad por suicidio respecto a los regresores edad, periodo y cohorte. En este aspecto, en relación con la variable edad, las tasas de mortalidad por suicidio presentan una tendencia creciente entre los 15 y los 49 años, edad a partir de la cual se produce una disminución de las ratios de mortalidad por suicidio. Finalmente, dichas tasas vuelven a experimentar un importante crecimiento a partir de los 70 años. Por otra parte, si se considera el regresor periodo, las tasas de mortalidad por suicidio poseen una tendencia decreciente entre 1985 y 2009; presentando, a partir del año 2010, una tendencia creciente. Finalmente, analizando el regresor cohorte, las tasas de mortalidad por suicidio poseen una tendencia creciente en relación con las cohortes nacidas entre 1900 y 1965: poseyendo, a partir de dicha cohorte, una tendencia decreciente.

Por otro lado, la figura 5.2 refleja el gráfico *sparsity*. En este aspecto, dicha figura permite representar las observaciones con valores inferiores a dos y uno, como se observa en la parte superior de la imagen. En este caso, las gráficas se encuentran vacías; ya que, el número de fallecimientos por suicidio y la población, no presentan valores inferiores a estos puntos de corte.

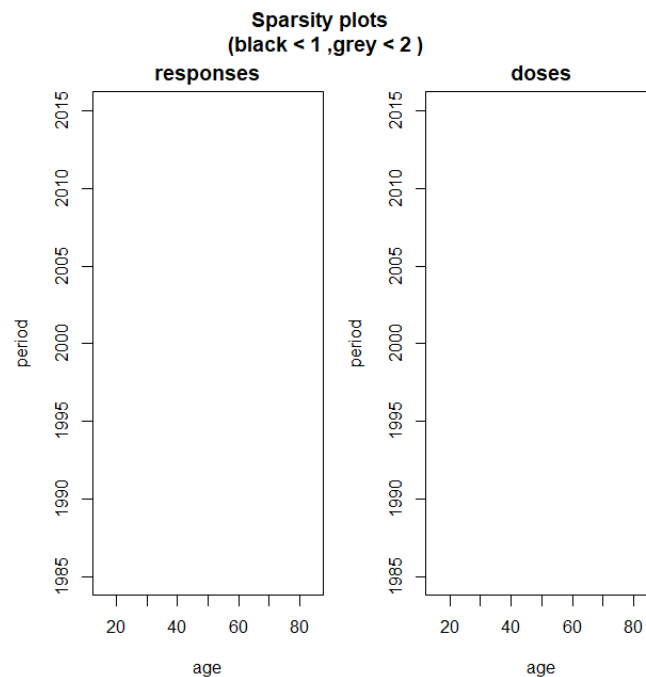


Figura 5.2: Gráfico *sparsity* para los suicidios (*responses*) y la población (*doses*).

Por otro lado, a continuación, se introducirán las gráficas para las muertes por suicidio (*responses*), la población (*doses*) y las tasas de mortalidad por suicidio (*rates*); usando, para ello, únicamente, dos índices temporales.

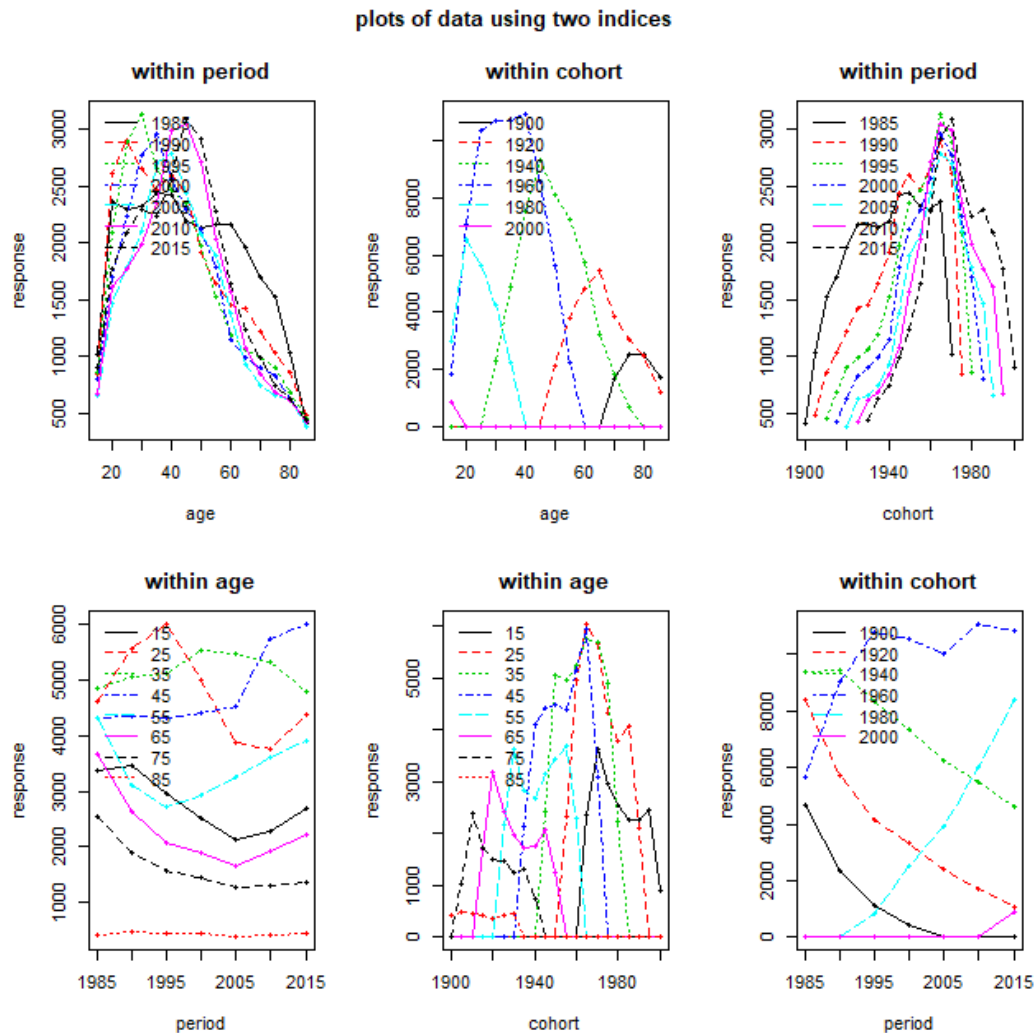


Figura 5.3: Representación de las cifras de suicidios totales (*responses*) teniendo en cuenta, únicamente, dos índices temporales.

De tal forma, en primer lugar, se introducen, en la figura 5.3, las gráficas referentes a las cifras de fallecidos por suicidio (*responses*). En este aspecto, en la primera imagen, se muestra la relación gráfica existente entre las variables edad y periodo; concentrándose, las cifras de suicidios, con independencia del periodo, entre los 30 y los 50 años. A partir de esta edad, el volumen de suicidios se ve reducido, independientemente del periodo. Por otra parte, en la segunda imagen, aparecen representadas las variables edad y cohorte; de forma que, a medida que la cohorte de nacimiento adquiere mayor longevidad, la edad en la que se produce el suicidio es mayor. Por otro lado, en la tercera imagen, se representan los regresores cohorte y periodo. En este aspecto, en los periodos acaecidos entre 1990 y 2015, la mayoría de suicidios fueron cometidos por personas nacidas entre 1960 y 1980. Por otra parte, en relación con la cuarta imagen, se observa que, el grupo de edad que posee entre 85 y 89 años, posee una tendencia estable, en cuanto al número de suicidios, durante el periodo analizado. Por otro lado, el grupo de individuos más joven (los que poseen entre 15 y 25 años) ha experimentado, a partir del año 2005, un incremento en el número de suicidios; al igual que, los grupos de edad que poseen entre 25- 34 años, 45- 54 años y 65- 74 años. Sin embargo, los suicidios cometidos

por personas con edades comprendidas entre los 75 y los 84 años se han reducido, entre 1985 y 2019, de forma importante. Por otro lado, en la quinta imagen están representadas las variables cohorte (eje OX) y edad (eje OY); mostrando información similar, por tanto, a la contenida en la segunda imagen. Finalmente, en la sexta gráfica se observa que, las cohortes nacidas en las décadas de los sesenta y setenta del siglo XX, presentan las mayores cifras de fallecimientos por suicidio; mientras que, las cohortes de nacimiento más antiguas, poseen una tendencia decreciente, a medida que avanza el tiempo. Esta circunstancia es lógica; ya que, las mayores cifras de suicidio se dan en la edad adulta (principalmente entre los 40 y 60 años); por lo que, las cohortes de nacimiento con esas edades, presentan mayores cifras de suicidio.

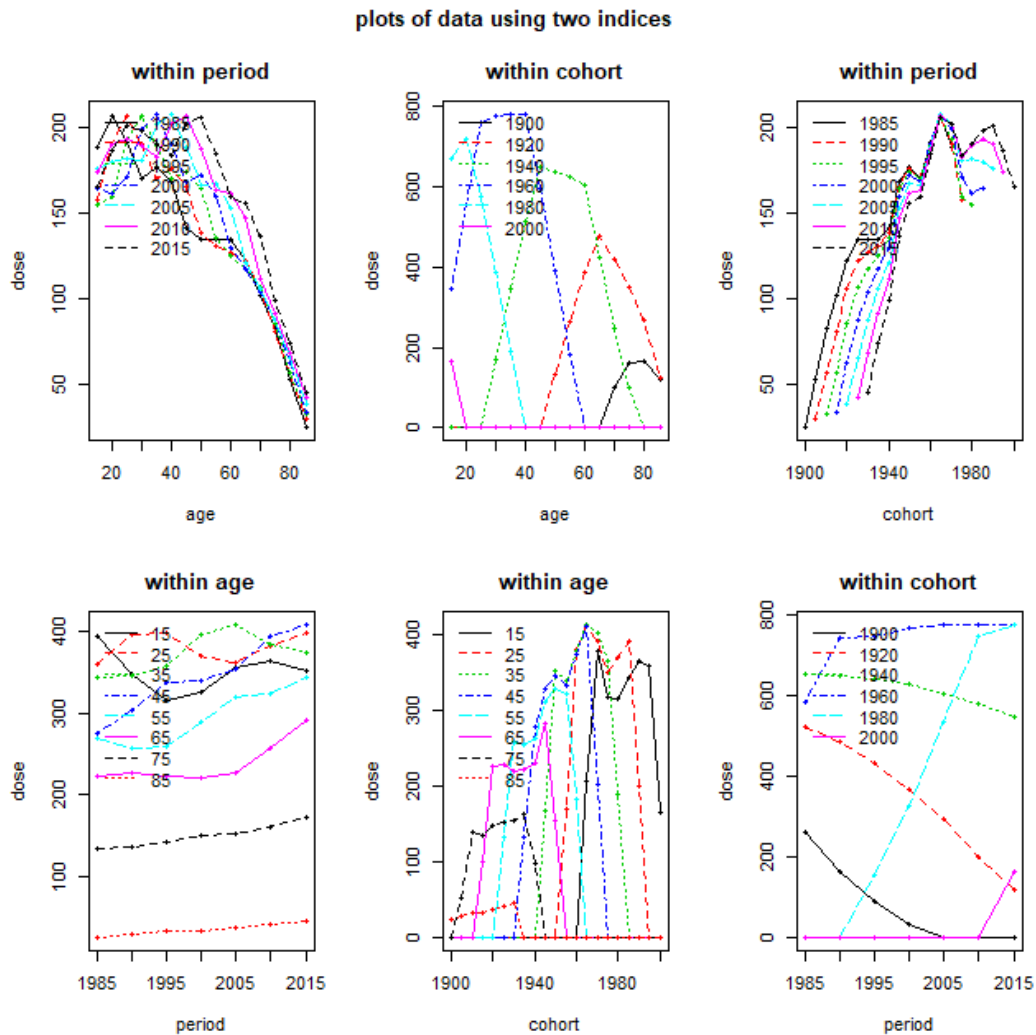


Figura 5.4: Representación de la población (*dose*) teniendo en cuenta, únicamente, dos índices temporales.

Por otra parte, en la figura 5.4 se introducen las gráficas referentes a la población (*dose*). De tal forma, la primera imagen muestra la relación existente entre las variables periodo y edad. En este aspecto, los máximos poblacionales, a medida que avanza la variable periodo, tienden a poseer una mayor edad; mostrando, este hecho, un progresivo envejecimiento de la población inglesa y galesa. Por otro lado, la segunda imagen representa las variables edad y cohorte; mostrando, en este aspecto, un decrecimiento de la población nacida en cada cohorte, a medida que aumenta la edad de dichas personas. Respecto a la tercera imagen, esta muestra la relación existente entre las variables periodo y cohorte. De tal forma, las cohortes más antiguas, a medida que transcurre el tiempo (aumento del periodo) pierden relevancia en cuanto al total poblacional. Por otro lado, la cuarta imagen refleja la estructura de la pirámide poblacional inglesa y galesa; ya que, las

generaciones nacidas durante el *Baby Boom* británico, tienden a acumular la mayor parte de la población total. Además, durante el periodo de análisis, se observa una tendencia creciente en relación con el número de personas que viven más de 65 años (envejecimiento progresivo de la población); mientras que, la población total con menos de 25 años, se ha visto sometida, durante los últimos periodos de la serie, a una tendencia decreciente. Por otro lado, en la imagen 5 se muestra la evolución de la variable edad en función de la cohorte de nacimiento; por lo que, las cohortes más jóvenes, son las que poseen una mayor población de reducida edad. Finalmente, la sexta imagen muestra la relación gráfica existente entre las variables periodo y cohorte. De tal forma, a medida que aumenta la variable periodo, las cohortes de nacimiento más recientes van ganando protagonismo, en relación con el total poblacional, y en detrimento de las cohortes más antiguas.

Finalmente, en la figura 5.5 se muestran las gráficas referentes a las tasas de mortalidad por suicidio. Estas ratios serán estudiadas, en profundidad, en la sección 5.2.

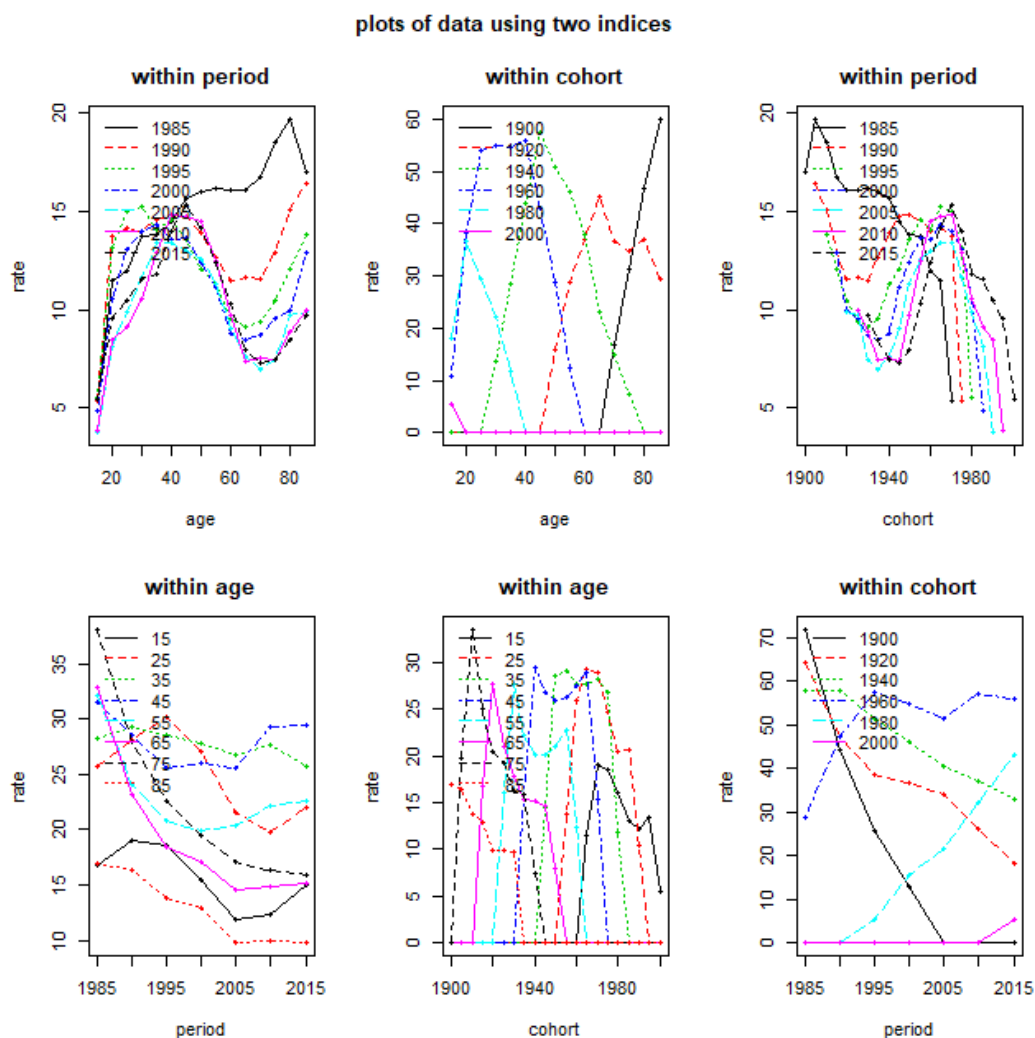


Figura 5.5: Representación de las tasas de mortalidad por suicidio (*rates*) teniendo en cuenta, únicamente, dos índices temporales.

Tras realizar el correspondiente análisis gráfico de la base de datos a analizar, se procederá con la estimación del modelo.

5.1.3. Estimación del modelo

En este epígrafe, se procederá con el ajuste del modelo; considerando, para ello, la familia *poisson. dose. response*; ya que, se utilizarán tanto las muertes totales por suicidio (*response*) como, la población total (*dose*). De tal forma, para proceder con la estimación del modelo edad- periodo- cohorte correspondiente, se empleará la función *apc.fit.model*:

```
> modelo_numero_suicidios_totales=apc.fit.model(numero_suicidios_totales, model.family =
"poisson.dose.response",model.design = "APC")
```

Una vez aplicada esta función, se obtendrán los coeficientes canónicos del modelo:

```
> modelo_numero_suicidios_totales$coefficients.canonical
##              Estimate Std. Error      z value      Pr(>|z|)
## level          2.723828655 0.01431871 190.2286673 0.000000e+00
## age slope      -0.142640923 0.01481826  -9.6260252 6.208470e-22
## cohort slope  -0.093827329 0.01763016  -5.3219795 1.026441e-07
## DD_age_25     -0.695660784 0.02394968 -29.0467668 1.690513e-185
## DD_age_30     -0.031846230 0.01990759  -1.5997032 1.096644e-01
## DD_age_35     -0.030646974 0.01897411  -1.6151999 1.062674e-01
## DD_age_40      0.015717574 0.01856306   0.8467125 3.971554e-01
## DD_age_45     -0.026270731 0.01834791  -1.4318102 1.521981e-01
## DD_age_50     -0.055733849 0.01868829  -2.9822872 2.861035e-03
## DD_age_55     -0.063012678 0.01983590  -3.1766981 1.489620e-03
## DD_age_60     -0.103912107 0.02174780  -4.7780513 1.770022e-06
## DD_age_65      0.037181063 0.02429031   1.5306954 1.258447e-01
## DD_age_70      0.062481039 0.02680105   2.3312911 1.973802e-02
## DD_age_75      0.059625842 0.02921724   2.0407761 4.127308e-02
## DD_age_80      0.044458257 0.03194506   1.3917098 1.640103e-01
## DD_age_85     -0.018504431 0.03708419  -0.4989844 6.177904e-01
## DD_period_1995 0.008600707 0.01506717   0.5708243 5.681188e-01
## DD_period_2000 -0.006992607 0.01543902  -0.4529179 6.506079e-01
## DD_period_2005 -0.037621153 0.01581383  -2.3790030 1.735953e-02
## DD_period_2010 0.139049370 0.01613407   8.6183713 6.791314e-18
## DD_period_2015 -0.021101780 0.01579091  -1.3363244 1.814433e-01
## DD_cohort_1910 -0.195215640 0.07612718  -2.5643355 1.033736e-02
## DD_cohort_1915 -0.048687334 0.04959867  -0.9816258 3.262842e-01
## DD_cohort_1920 -0.015437242 0.04067137  -0.3795604 7.042717e-01
## DD_cohort_1925 0.052397378 0.03525189   1.4863706 1.371811e-01
## DD_cohort_1930 -0.046481436 0.03130981  -1.4845647 1.376592e-01
## DD_cohort_1935 0.041119216 0.02908558   1.4137322 1.574406e-01
## DD_cohort_1940 0.071726015 0.02752064   2.6062627 9.153623e-03
## DD_cohort_1945 0.022869671 0.02544580   0.8987601 3.687805e-01
## DD_cohort_1950 0.057629651 0.02272401   2.5360686 1.121048e-02
## DD_cohort_1955 0.020876603 0.02087310   1.0001678 3.172293e-01
## DD_cohort_1960 -0.053522823 0.01984294  -2.6973233 6.989936e-03
## DD_cohort_1965 0.060416478 0.01864932   3.2396081 1.196941e-03
## DD_cohort_1970 -0.012722666 0.01779167  -0.7150911 4.745528e-01
## DD_cohort_1975 -0.111836980 0.01892754  -5.9086917 3.448353e-09
## DD_cohort_1980 -0.070863495 0.02251524  -3.1473563 1.647541e-03
## DD_cohort_1985 0.071551366 0.02678172   2.6716497 7.547939e-03
## DD_cohort_1990 0.039620882 0.03157163   1.2549519 2.094962e-01
## DD_cohort_1995 0.040076085 0.03904923   1.0262963 3.047519e-01
## DD_cohort_2000 0.241414769 0.05605214   4.3069678 1.655077e-05
```

Por tanto, en base a esta salida, se pueden estudiar las tres componentes asociadas al plano lineal del modelo edad- periodo- cohorte; obteniéndose, estas, mediante la imposición de restricciones que fijan las segundas diferencias a cero. En este aspecto, el nivel sería igual a $\hat{v}_0 = 2,723828655$; mientras que, las pendientes

del regresor edad y cohorte serían iguales, respectivamente, a $\hat{v}_{\text{age}} = -0,142640923$ y $\hat{v}_{\text{cohorte}} = -0,093827329$. Por otro lado, considerando un nivel de significación del 5% se tendría que, las pendientes de las variables edad y cohorte, resultan ser significativas.

Tras la obtención y el posterior análisis de los coeficientes de correlación canónicos asociados al modelo general, se procederá con el análisis gráfico de dicho modelo. Para ello, se empleará el siguiente comando:

```
> apc.plot.fit.all(modelo_numero_suicidios_totales)
```

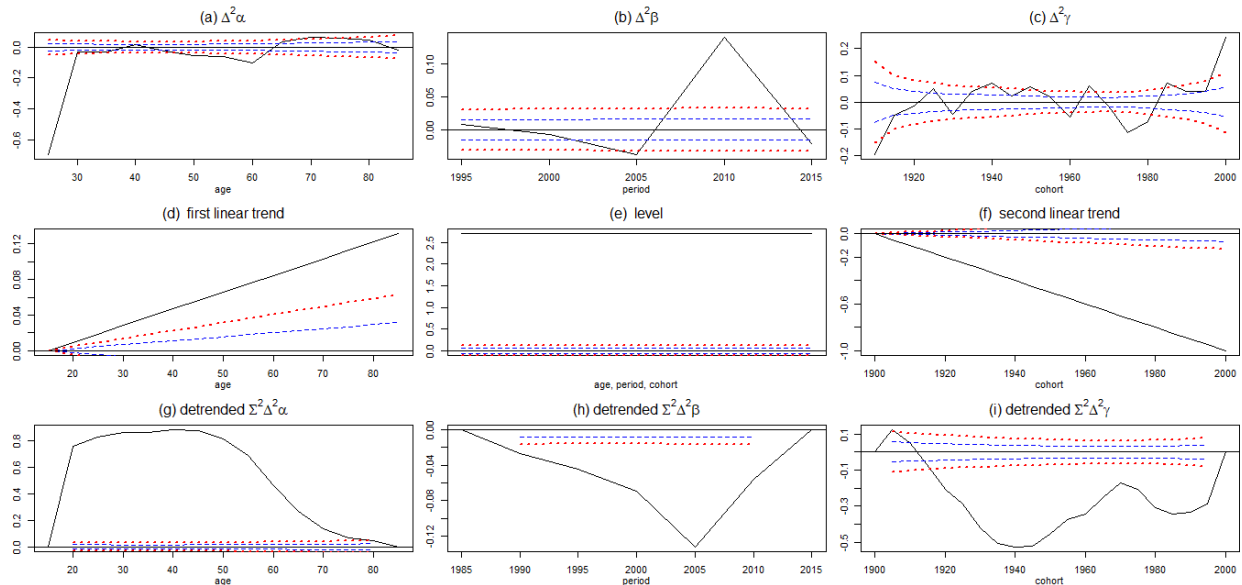


Figura 5.6: Representación de las estimaciones de los parámetros haciendo uso de la metodología *detrend*.

En primer lugar, en la figura 5.6 se muestran los parámetros canónicos asociados al modelo edad- periodo-cohorte, junto a la representación *detrended*. En este aspecto, los tres primeros gráficos muestran las estimaciones de los parámetros canónicos correspondientes a las segundas diferencias; que se encuentran relacionadas, respectivamente, con las variables edad, periodo y cohorte. Además, dichas estimaciones son representadas junto a una banda de confianza centrada en cero. Por otro lado, las gráficas (d), (e) y (f) reflejan el plano lineal del modelo, cuyos parámetros fueron introducidos anteriormente. Finalmente, las gráficas contenidas en la tercera fila de la figura 5.6, permiten obtener la parte no lineal asociada a los regresores temporales del modelo (efectos de las variables identificados *ad hoc*). De tal forma, al utilizar el gráfico tipo *detrend*, se consiguen mostrar las dobles sumas de las dobles diferencias (tipo *detrend*); por lo que, estas comienzan y concluyen en cero.

El estudio de estos últimos gráficos puede arrojar cuestiones interesantes. En primer lugar, como se señaló anteriormente, las imágenes (a), (b) y (c) muestran los parámetros estimados correspondientes a las segundas diferencias ($\Delta^2\alpha_i$, $\Delta^2\beta_j$ y $\Delta^2\gamma_k$, respectivamente). En este aspecto, no se aprecia ruido alguno en los gráficos (a) y (c); por lo que, no será necesario realizar un análisis de submuestras. Por otro lado, en las imágenes (d), (e) y (f) se muestran el nivel y las pendientes estimadas; cuyos valores fueron introducidos, anteriormente, al obtener los coeficientes canónicos del modelo. Finalmente, las subgráficas (g), (h) e (i) muestran las dobles sumas de las dobles diferencias utilizando el método *detrended*; por lo que, estas comienzan y acaban en cero. En este aspecto, las dobles sumas del regresor edad generan un gráfico cóncavo; mientras que, las partes no lineales asociadas al regresor periodo, muestran un aumento en las cifras de suicidios a partir del periodo 2005- 2009. Finalmente, en relación con la variable cohorte, se observan ciertas discontinuidades; motivadas, estas últimas, por los periodos asociados a las guerras mundiales y las etapas de entreguerras acaecidos en la primera mitad del siglo XX. Sin embargo, las cohortes nacidas en etapas recientes, poseen cifras de suicidio superiores, en comparación con las generaciones nacidas en la segunda mitad del siglo XX.

Por otro lado, partiendo de las conclusiones planteadas por Nielsen (2015) [129], el predictor lineal (gráficos (d), (e) y (f) de la figura 5.6) asociado a un individuo nacido, por ejemplo, en 1955 y fallecido, por suicidio, en 1995 (con 40 años de edad), podría ser calculado a partir de la suma de los siguientes elementos (que conforman los gráficos que se acaban de presentar; justificando, por tanto, su importancia):

- La tendencia del regresor edad concerniente a los 40 años (*gráfico d*).
- El nivel representado en el *gráfico e*.
- La tendencia, asociada a la variable cohorte, y evaluada, a su vez, en el año de nacimiento del sujeto (en este caso 1955); correspondiéndose, pues, con el *gráfico f*.
- El efecto del regresor edad, sin tener en cuenta la tendencia; y considerando, a su vez, la edad del sujeto (40 años, en este caso). Este elemento se encuentra representado en el *gráfico g*.
- El efecto asociado al regresor periodo (sin tener en cuenta la tendencia); considerando, a su vez, el año de fallecimiento por suicidio (1995, en este ejemplo). Se consideraría en este caso, por tanto, la *gráfica h*.
- El efecto referente al regresor cohorte, sin considerar la tendencia; siguiendo, pues, los efectos representados en el *gráfico i*. De tal forma, en este hipotético caso se evaluaría, el efecto del regresor cohorte, en el año 1955 (nacimiento del individuo).

Por otro lado, la figura 5.7 permite identificar si las cifras totales de suicidio (*responses*) son relevantes a la hora de construir el modelo edad- periodo- cohorte; mostrando la existencia, además, de valores atípicos (*outliers*). En este aspecto, si una cifra de suicidios posee la forma de un círculo negro, ésta se encontrará centrada respecto al resto de respuestas; formando parte, pues, del 80% central de la distribución. Por otra parte, los triángulos hacen referencia a las respuestas que se encuentran en las colas de la distribución. En este caso, teniendo en cuenta los datos analizados, la mayor parte de las respuestas estarían situadas en las colas de la distribución, al tener asociadas, en la gráfica, formas geométricas triangulares. En este aspecto, en base a las características de dicha figura; podría señalarse que, las cifras de suicidio, presentan importantes variaciones durante el periodo de análisis.

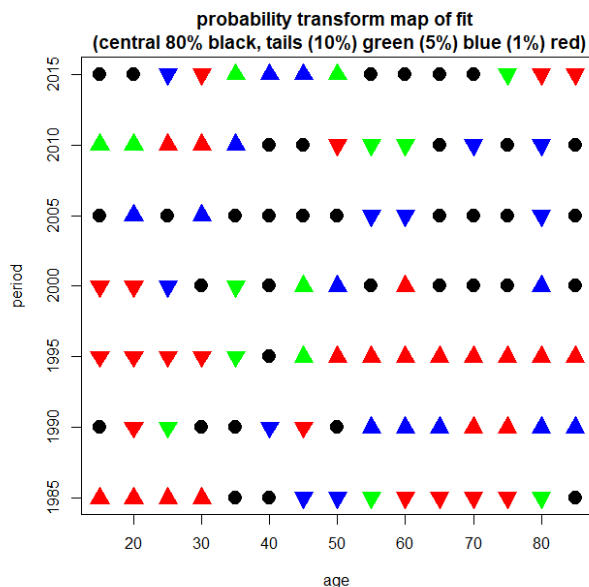


Figura 5.7: Mapa de ajuste de la probabilidad.

Por otro lado, en la figura 5.8 se muestran los residuos obtenidos al ajustar el modelo. En este aspecto, los residuos con valores iguales o cercanos a cero, reflejan un correcto ajuste del modelo. Por tanto, la mayoría de

los valores se encuentran ajustados de forma correcta; aunque, en algunas situaciones, el modelo no presenta un ajuste eficiente. Estos valores se encuentran, principalmente, en los grupos de edad extremos (15- 19 años y 85- 89 años); pudiendo estar motivado, este hecho, por las menores cifras de suicidio que poseen dichos grupos de edad.

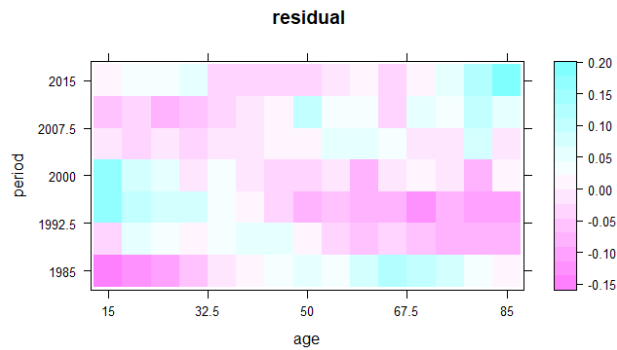


Figura 5.8: Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las cifras de suicidio.

Por otra parte, en la figura 5.9 se muestra la distribución gráfica de los valores ajustados obtenidos en base al modelo general. En este aspecto, la mayoría de los valores ajustados poseen cifras elevadas; mientras que, los valores más bajos, se encuentran en los extremos de la distribución, concretamente en el grupo de edad más joven (15- 19 años) y en el intervalo con mayor edad (85- 89 años).

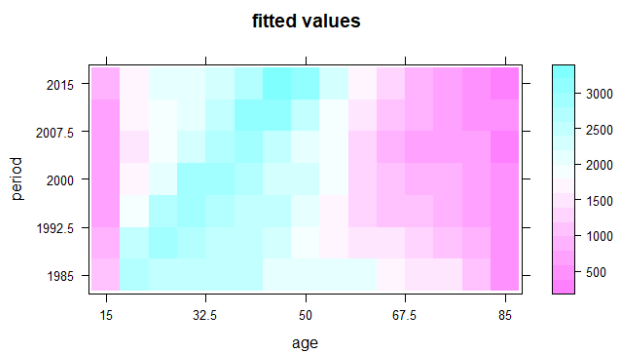


Figura 5.9: Representación gráfica de los valores ajustados obtenidos al implementar un modelo edad- periodo- cohorte aplicado a las cifras de suicidio.

Por otro lado, en la figura 5.10 se introducen los valores del predictor lineal relacionados con cada observación. En este aspecto, la mayoría de predictores lineales poseen valores elevados; correspondiéndose, estos, con los elementos centrales de la distribución. Sin embargo, los valores con predictores lineales más pequeños, se encuentran situados en el extremo superior de la gráfica; coincidiendo, por tanto, con el grupo de población con mayor edad (85- 89 años).

En la figura 5.11, se muestra la distribución del número de suicidios totales acaecidos durante el periodo de análisis. En este aspecto, la mayoría de suicidios se producen en la edad adulta (en concreto, entre los 20 y los 69 años); mientras que, los suicidios que se producen en los grupos de edad comprendidos entre los 15- 19 años y 70- 89 años, presentan cuantías inferiores. En este aspecto, dicha imagen posee un notable parecido en relación con la figura 5.9; por lo que, el ajuste realizado por el modelo edad- periodo- cohorte, puede ser apropiado.

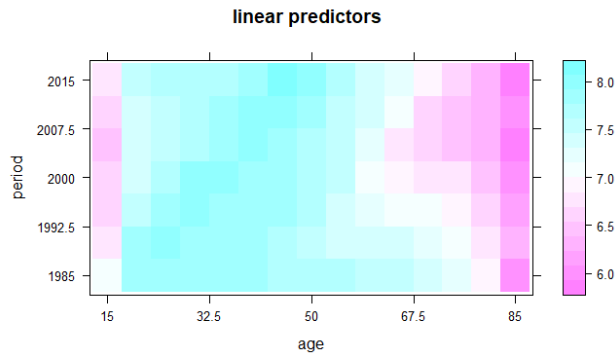


Figura 5.10: Representación gráfica de los predictores lineales obtenidos al implementar un modelo edad-periodo- cohorte a las cifras de suicidio.

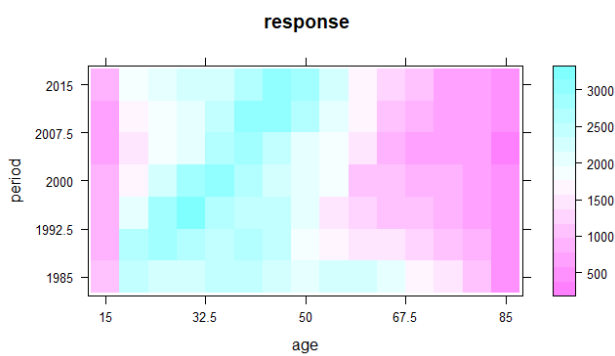


Figura 5.11: Representación gráfica de las respuestas (*responses*).

Finalmente, en la figura 5.12 se muestra la distribución asociada a las tasas de mortalidad por suicidio. De tal forma, las menores ratios de suicidio se concentran en los grupos de edad más jóvenes (15- 19 años).

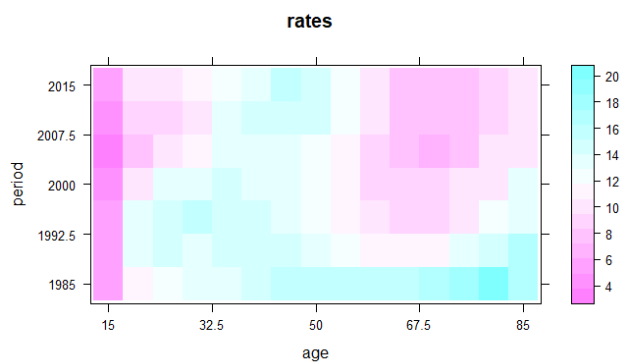


Figura 5.12: Representación gráfica de las tasas de mortalidad por suicidio (*rates*).

5.1.4. Inferencia sobre el modelo

Una vez estudiadas las salidas obtenidas al estimar el modelo edad- periodo- cohorte, se obtendrá la *tabla deviance*. En este aspecto, el objetivo de este epígrafe consiste en averiguar si, un determinado submodelo, permite ajustar, en mayor medida que el modelo edad- periodo- cohorte obtenido anteriormente, las cifras de suicidio analizadas.

Para obtener la *tabla deviance* se debe ejecutar el siguiente comando:

```
> apc.fit.table(numero_suicidios_totales,"poisson.dose.response")
##      deviance df.residual prob(>chi_sq) LR vs.APC df vs.APC prob(>chi_sq)
## APC    578.190         65           0      NaN      NaN      NaN
## AP   2807.401         84           0  2229.211      19         0
## AC    849.620         70           0   271.430       5         0
## PC   7790.802         78           0  7212.613      13         0
## Ad   3194.108         89           0  2615.918      24         0
## Pd  12108.262         97           0 11530.073      32         0
## Cd   8243.018         83           0  7664.828      18         0
## A    5016.468         90           0  4438.278      25         0
## P   12325.159         98           0 11746.969      33         0
## C    8525.596         84           0  7947.406      19         0
## t   12445.394        102           0 11867.204      37         0
## tA  14153.290        103           0 13575.100      38         0
## tP  12664.749        103           0 12086.560      38         0
## tC  13357.622        103           0 12779.432      38         0
## 1   14326.008        104           0 13747.818      39         0
##      aic
## APC  1615.987
## AP   3807.198
## AC   1877.416
## PC   8802.599
## Ad   4183.904
## Pd  13082.059
## Cd   9244.815
## A    6004.265
## P   13296.956
## C    9525.393
## t   13409.191
## tA  15115.087
## tP  13626.546
## tC  14319.419
## 1   15285.805
```

Si se analizan las columnas de *deviance* (primera columna) y *aic* (última columna), el modelo edad-periodo-cohorte es el que presenta los menores valores, en relación con ambos indicadores; por lo que, los efectos de los regresores edad, periodo y cohorte, resultan ser significativos a la hora de analizar las cifras brutas de suicidio. Por tanto, el modelo edad-periodo-cohorte es el que presenta un mayor ajuste, en comparación con el resto de submodelos.

5.2. Aplicación de un *APC model* a las tasas de mortalidad por suicidio

Tras realizar el análisis de las cifras brutas de suicidios concernientes al periodo 1985- 2019 (considerando como *doses* al total poblacional), se procederá con el estudio de las tasas de mortalidad por suicidio (por 100000 habitantes) para el citado periodo. De tal forma, en este caso, se debe utilizar un modelo *gaussian.response*; ya que, únicamente, se emplearán las tasas de mortalidad por suicidio.

Por otro lado, la mayoría de los estudios que analizan los efectos de las variables edad, periodo y cohorte en materia de suicidios; utilizan, como respuestas, las tasas de mortalidad por suicidio, como bien se explicó en el epígrafe de revisión de la literatura (consultar sección 2.2). Por tanto, se compararán los resultados

obtenidos en los posteriores epígrafes con los encontrados, anteriormente, por otros autores.

5.2.1. Creación del objeto de datos

En primer lugar, se deben introducir los datos que se utilizarán en el estudio en el software R. Para ello, se ha utilizado la función *read_excel*; que se encuentra disponible, a su vez, en la librería *readxl*; ya que, los datos del estudio se encontraban contenidos en un archivo *.xls*. Tras introducir los datos, estos han sido depurados y transformados en matrices; siguiendo, para ello, los pasos contenidos en el anexo de este documento (apéndice A.1.2)

De tal forma, una vez se dispone de la base de datos a analizar, se debe crear el correspondiente objeto de tipología *apc.data.list*. Para ello, se deben ejecutar los siguientes comandos:

```
> library(apc) # Se carga la librería
# Se debe transformar el dataframe de las tasas en una matriz
# (requisito para trabajar con el paquete apc)
> tasa_suicidios_totales_matrix=as.matrix(t_Tasas_Suicidios_Totales)
# Se crea el objeto con la función apc.data.list
> tasa_suicidios_totales=apc.data.list(tasa_suicidios_totales_matrix,"AP",
age1=15, per1 = 1985, unit = 5, label=
"Suicide mortality rates England and Wales (period 1985- 2019)")
> tasa_suicidios_totales
## $response
##      1985-1989 1990-1994 1995-1999 2000-2004 2005-2009 2010-2014 2015-2019
## 15-19      5.36      5.38      5.50      4.88      3.74      3.84      5.46
## 20-24     11.44     13.70     13.12     10.54     8.16      8.48     9.52
## 25-29     11.98     14.12     14.96     13.06     9.86      9.16    10.44
## 30-34     13.72     13.98     15.22     13.98     11.66     10.54    11.54
## 35-39     13.82     14.58     14.04     14.30     13.42     12.88    11.76
## 40-44     14.46     14.78     14.58     13.56     13.40     14.84    13.98
## 45-49     15.62     14.72     13.52     13.64     13.00     14.76    15.28
## 50-54     15.96     13.88     12.08     12.36     12.56     14.50    14.16
## 55-59     16.12     12.62     11.34     11.16     11.28     12.48    12.38
## 60-64     16.06     11.50      9.54      8.80      9.02      9.68    10.28
## 65-69     16.08     11.62      9.14      8.44      7.64      7.38     7.96
## 70-74     16.76     11.52      9.34      8.72      6.98      7.54     7.26
## 75-79     18.50     12.92     10.50      9.52      7.44      7.46     7.48
## 80-84     19.68     15.06     12.08      9.96      9.68      8.84     8.42
## 85-89     16.96     16.36     13.80     12.92      9.84      9.92     9.74
##
## $dose
## NULL
##
## $data.format
## [1] "AP"
##
## $age1
## [1] 15
##
## $per1
## [1] 1985
##
## $coh1
## [1] 5
##
```

```
## $unit
## [1] 5
##
## $per.zero
## NULL
##
## $per.max
## NULL
##
## $time.adjust
## [1] 0
##
## $label
## [1] "Suicide mortality rates England and Wales (period 1985- 2019)"
##
## $n.decimal
## NULL
```

Por tanto, se han seguido los mismos pasos que se realizaron en el epígrafe 5.1.1; aunque en este caso, únicamente, se trabajará con las tasas de mortalidad por suicidio.

5.2.2. Exploración de los datos

De tal forma, tras introducir el objeto en formato *apc.data.list*, se procederá con el análisis gráfico de las tasas de mortalidad por suicidio de Inglaterra y Gales, por 100000 habitantes, pertenecientes al periodo 1985- 2019.

Para realizar el análisis gráfico, se debe implementar el siguiente comando:

```
> apc.plot.data.all(tasa_suicidios_totales)
```

En primer lugar, la figura 5.13 permite mostrar la suma de los datos; considerando, individualmente, los efectos temporales asociados a las variables edad, periodo y cohorte.

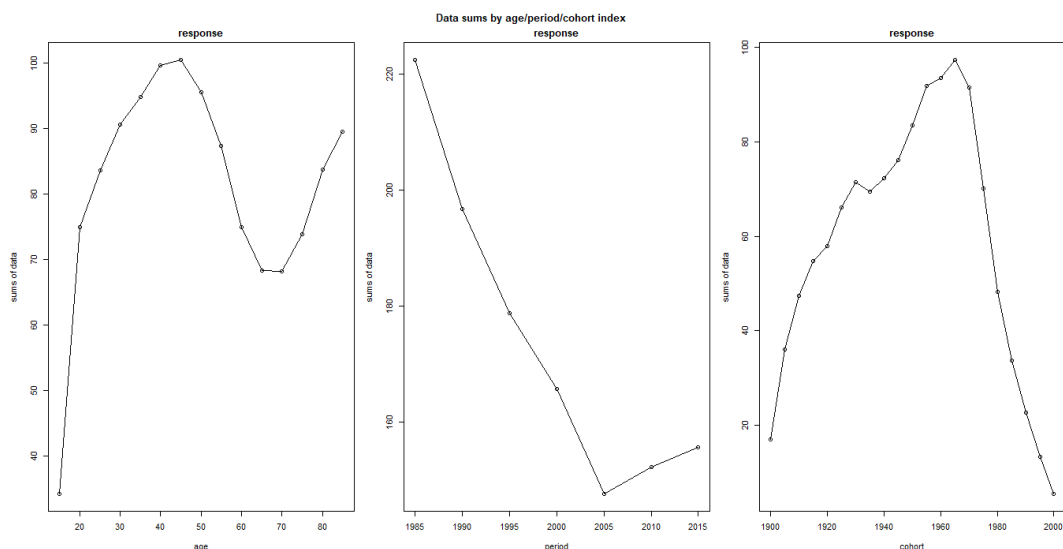


Figura 5.13: Suma de los datos en función de las variables edad, periodo y cohorte.

De tal forma, en relación con la variable edad, se observa un aumento de las tasas de mortalidad por suicidio entre los 15 y los 49 años. Posteriormente, entre los 50 y los 74 años se produce una reducción, en el periodo analizado, de las ratios de mortalidad por suicidio. Finalmente, entre los 75 y los 89 años, se vuelve a producir un aumento de las tasas de mortalidad por suicidio.

Por otro lado, si se analiza el regresor periodo, se observa una disminución de las tasas de mortalidad por suicidio entre 1985 y 2009. Finalmente, entre 2010 y 2019, se produce un incremento de las ratios de mortalidad por suicidio.

Por otra parte, en cuando a la variable cohorte, las cohortes nacidas antes del año 1966, han incrementado las tasas de mortalidad por suicidio; mientras que, las cohortes nacidas a partir de 1966, han reducido dichas ratios. Por tanto, desde un punto de vista visual, existe un importante efecto asociado a la variable cohorte.

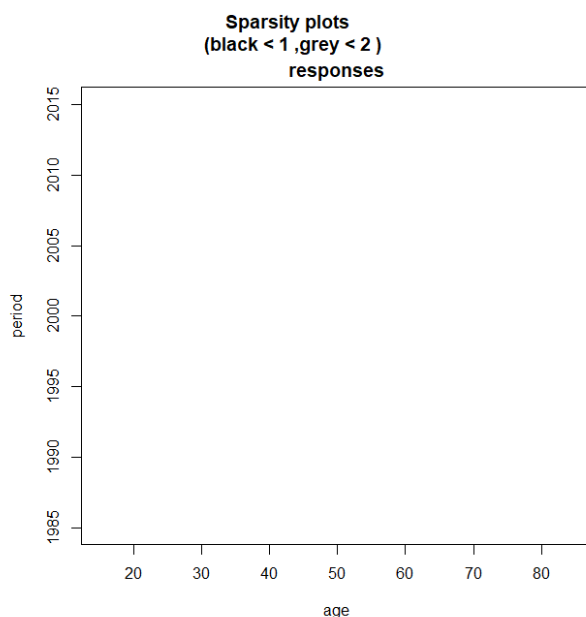


Figura 5.14: Gráfico tipo *sparsity* para las tasas de mortalidad por suicidio.

Por otro lado, en la figura 5.14 se muestra el diagrama tipo *sparsity*, que permite mostrar la dispersión de los datos. De tal forma, esta figura muestra las observaciones que poseen los valores más reducidos de la base de datos; estando, en este caso, vacía.

Por otra parte, en la figura 5.15 se introduce la evolución de las tasas de mortalidad por suicidio; en función, únicamente, de dos de los tres regresores temporales considerados en el análisis. En este aspecto, autores como Fannon y Nielsen (2020) [8] señalan que si, al eliminar un regresor, las líneas de la gráfica obtenida resultan ser paralelas, dicha variable podría ser eliminada del modelo. Sin embargo, se debe ser cauto con esta afirmación; siendo necesario, por tanto, la implementación de pruebas analíticas que confirmen esta idea. De tal forma, las gráficas que han sido obtenidas serán interpretadas a continuación:

- **Primera gráfica.** En este caso, se representa la evolución de las tasas de mortalidad por suicidio en función de las variables edad y periodo. De tal forma, las personas menores de 50 años han mantenido, durante todo el periodo de análisis, tasas de mortalidad por suicidio similares. Sin embargo, las ratios de mortalidad por suicidio de las personas mayores de 50 años, poseen una tendencia decreciente. De tal forma, a pesar de esta última circunstancia, la gráfica de las tasas de mortalidad por suicidio, en función del periodo, presenta cierto paralelismo.
- **Segunda gráfica.** Esta imagen muestra la evolución de las tasas de mortalidad por suicidio en relación con los regresores edad y cohorte. De tal forma, como es lógico, las cohortes nacidas en los años 1900

plots of data using two indices

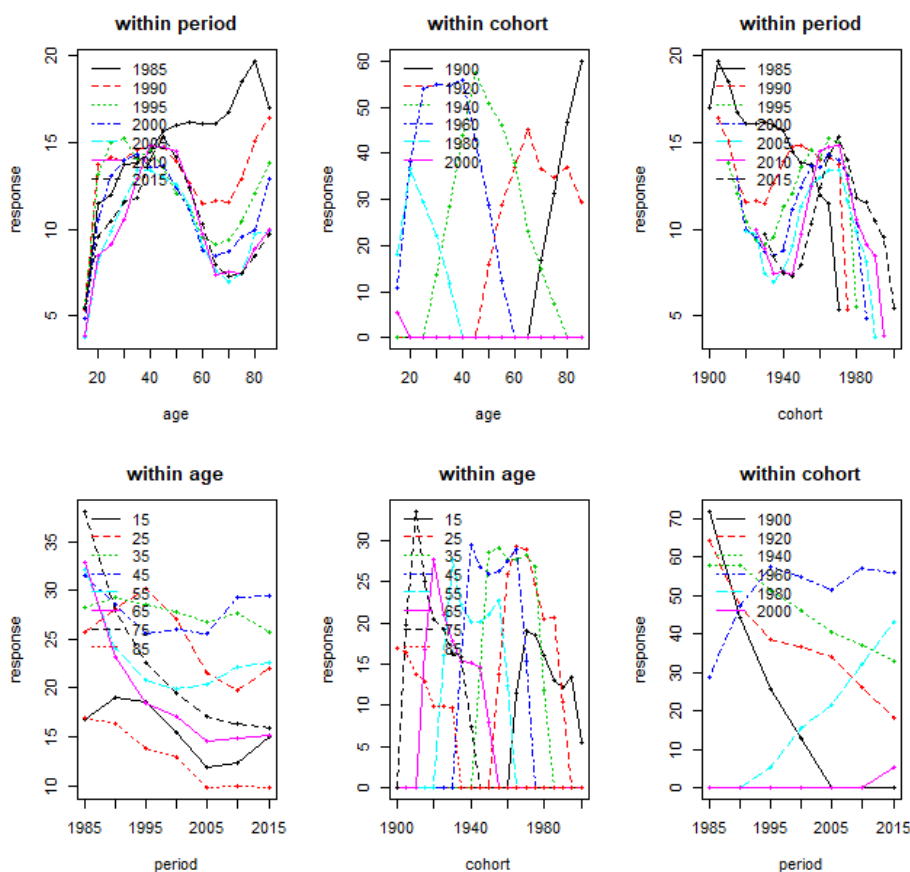


Figura 5.15: Gráfica de las tasas de mortalidad por suicidio utilizando, únicamente, dos índices temporales.

y 2000, poseen las tasas máximas de suicidio, respectivamente, en los grupos con edades comprendidas entre los 15- 19 años y 85- 89. Por otro lado, en el resto de cohortes de nacimiento, se muestra un aumento de las tasas de mortalidad por suicidio hasta alcanzar los 50 años.

- **Tercera gráfica.** En este caso, se representan las tasas de mortalidad por suicidio en función de las variables periodo y cohorte. De tal forma, se observa una reducción de las tasas de mortalidad por suicidio a medida que se produce un avance en la variable cohorte.
- **Cuarta gráfica.** Esta imagen permite mostrar la evolución de las tasas de mortalidad por suicidio en función de los regresores edad y periodo. De tal forma, se observa una reducción de las ratios de mortalidad por suicidio durante el periodo de análisis (especialmente en los grupos que poseen edades comprendidas entre los 65- 89 años). Sin embargo, a partir del año 2010, se observa un aumento de las tasas de mortalidad por suicidio que afectan, principalmente, a las personas con edades comprendidas entre los 15- 34 años y 45- 54 años.
- **Quinta gráfica.** En este caso, se representan las tasas de mortalidad por suicidio en base a las variables edad y cohorte. De tal forma, se observa una disminución de las ratios de mortalidad por suicidio durante el periodo de análisis, con independencia del grupo de edad considerado; como se observa, igualmente, en la cuarta gráfica.
- **Sexta gráfica.** Esta imagen representa las tasas de mortalidad por suicidio en función de las variables periodo y cohorte. En este aspecto, a medida que pasa el tiempo (aumento de la variable periodo), las cohortes más antiguas van disminuyendo su importancia en relación con los suicidios. Por tanto, equivalentemente, las cohortes de nacimiento más jóvenes van adquiriendo una mayor importancia,

en relación con las tasas de mortalidad por suicidio. En este aspecto, dichas ratios experimentan un crecimiento entre los 15 y los 50 años, con independencia de la variable cohorte. De tal forma, a partir de los 75 años se vuelve a producir un incremento de las tasas de mortalidad por suicidio, con independencia de la cohorte de nacimiento del sujeto.

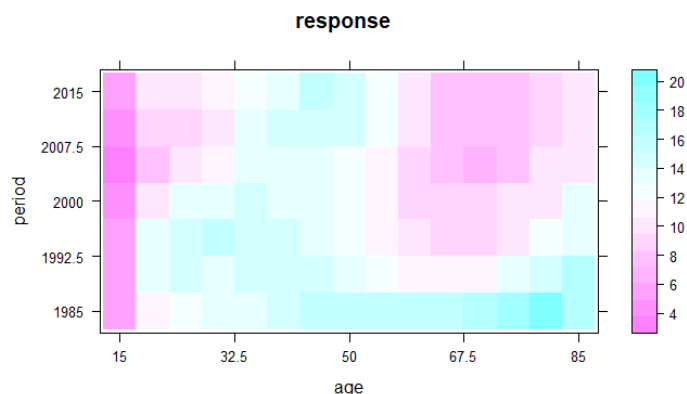


Figura 5.16: Distribución de las tasas de mortalidad por suicidio en función de las variables edad y periodo.

Por otro lado, en la figura 5.16 se muestran las tasas de mortalidad por suicidio en función de sus propios niveles. De tal forma, los colores más oscuros, como se observa en la leyenda de la figura, representan las tasas de mortalidad por suicidio más bajas; correspondiéndose, en este caso, con las ratios de mortalidad por suicidio de los jóvenes que poseen entre 15 y 19 años. Por otro lado, durante el periodo 1985- 1994, las tasas de mortalidad por suicidio más elevadas se concentraban en los grupos con mayores edades (70- 89 años); mientras que, durante el periodo que comprende a los años acaecidos entre 1995 y 2019, las ratios de mortalidad por suicidio con mayores cifras, se concentraron en el grupo de edad que poseía entre 35 y 49 años.

5.2.3. Estimación del modelo

Tras realizar el análisis gráfico, se procederá con la estimación del modelo edad- periodo- cohorte correspondiente. Para ello, se utilizará la función *apc.fit.modelo*; empleando, en este aspecto, la familia de modelos *gaussian.response* al estar utilizando, como respuestas (*responses*), las tasas de mortalidad por suicidio. En base a estas ideas, se aplicará el siguiente comando:

```
> modelo_tasa_suicidios_totales=apc.fit.modelo(tasa_suicidios_totales,
model.family = "gaussian.response",model.design = "APC")
```

Tras obtener el ajuste, se presentarán los valores asociados a los coeficientes canónicos del modelo general, con el objetivo de obtener los componentes asociados al plano lineal. De tal forma, los coeficientes canónicos se obtienen a través de la siguiente instrucción:

```
> modelo_tasa_suicidios_totales$coefficients.canonical
##           Estimate Std. Error   t value    Pr(>|t|)
## level      15.47443689  0.6086704 25.42334548 1.392130e-142
## age slope  -2.44319087  0.6816349 -3.58431025 3.379702e-04
## cohort slope -1.76330901  0.6820700 -2.58523184 9.731354e-03
## DD_age_25  -4.69403656  0.9823690 -4.77828264 1.767987e-06
## DD_age_30  -0.24800704  0.9739233 -0.25464739 7.989955e-01
## DD_age_35  -0.32583281  0.9708376 -0.33562030 7.371572e-01
## DD_age_40   0.20570947  0.9692155  0.21224327 8.319173e-01
## DD_age_45  -0.40740294  0.9682116 -0.42077884 6.739166e-01
## DD_age_50  -0.82989884  0.9675289 -0.85775094 3.910300e-01
## DD_age_55  -0.66676761  0.9673151 -0.68929724 4.906362e-01
```

## DD_age_60	-0.93675987	0.9675289	-0.96819831	3.329453e-01
## DD_age_65	0.71780226	0.9682116	0.74136924	4.584696e-01
## DD_age_70	0.54206268	0.9692155	0.55927983	5.759708e-01
## DD_age_75	0.50943194	0.9708376	0.52473444	5.997678e-01
## DD_age_80	0.37994379	0.9739233	0.39011673	6.964502e-01
## DD_age_85	-0.28800708	0.9823690	-0.29317608	7.693876e-01
## DD_period_1995	0.71353070	0.6634321	1.07551431	2.821445e-01
## DD_period_2000	0.24405216	0.6610475	0.36919006	7.119861e-01
## DD_period_2005	-0.48341807	0.6605803	-0.73180816	4.642857e-01
## DD_period_2010	1.37167215	0.6610475	2.07499787	3.798673e-02
## DD_period_2015	-0.23928038	0.6634321	-0.36067050	7.183458e-01
## DD_cohort_1910	-3.43517454	1.9359054	-1.77445373	7.598813e-02
## DD_cohort_1915	-0.07479262	1.5170369	-0.04930178	9.606788e-01
## DD_cohort_1920	-0.05511982	1.2967805	-0.04250513	9.660960e-01
## DD_cohort_1925	1.17348428	1.1528306	1.01791560	3.087181e-01
## DD_cohort_1930	-0.52636925	1.0487928	-0.50188106	6.157512e-01
## DD_cohort_1935	0.18226525	0.9797769	0.18602729	8.524234e-01
## DD_cohort_1940	0.67175627	0.9675315	0.69429914	4.874946e-01
## DD_cohort_1945	0.16916938	0.9673919	0.17487160	8.611805e-01
## DD_cohort_1950	0.65879034	0.9673530	0.68102373	4.958565e-01
## DD_cohort_1955	0.35646498	0.9673433	0.36849894	7.125012e-01
## DD_cohort_1960	-0.65669281	0.9673530	-0.67885541	4.972295e-01
## DD_cohort_1965	0.79556566	0.9673919	0.82238194	4.108596e-01
## DD_cohort_1970	-0.40768052	0.9675315	-0.42136151	6.734911e-01
## DD_cohort_1975	-1.08632513	0.9797769	-1.10874743	2.675392e-01
## DD_cohort_1980	-0.66830196	1.0487928	-0.63721066	5.239876e-01
## DD_cohort_1985	0.86700351	1.1528306	0.75206495	4.520120e-01
## DD_cohort_1990	0.53726090	1.2967805	0.41430366	6.786517e-01
## DD_cohort_1995	0.13852680	1.5170369	0.09131406	9.272430e-01
## DD_cohort_2000	1.29777231	1.9359054	0.67036970	5.026221e-01

En este aspecto, las tres primeras filas de la tabla de los coeficientes canónicos estimados, permiten definir el plano lineal sobre el que se asienta el modelo. En concreto, dicho plano lineal estaría definido por los tres primeros elementos de la columna *Estimate*; que poseen, a su vez, sus correspondientes niveles de significación en la columna $Pr(> |z|)$. De tal forma, el nivel sería igual a $\hat{v}_0 = 15,47443689$ ($Pr(> |z|) = 1,392130e - 142$); mientras que, las pendientes de los regresores edad y cohorte serían iguales, respectivamente, a $\hat{v}_{age} = -2,44319087$ ($Pr(> |z|) = 3,379702e - 04$) y $\hat{v}_{cohort} = -1,76330901$ ($Pr(> |z|) = 9,731354e - 03$). Por tanto, considerando un nivel de significación estándar del 5%, las pendientes de las variables edad y cohorte serían significativas.

Tras la obtención y el posterior análisis de los coeficientes de correlación canónicos asociados al modelo edad- periodo- cohorte, se procederá con el análisis gráfico asociado al mismo. Para ello, en primer lugar, se ejecutará el siguiente comando:

```
> apc.plot.fit.pt(modelo_tasa_suicidios_totales)
```

De tal forma, la figura 5.17 permite identificar si las tasas de mortalidad por suicidio resultan ser relevantes a la hora de construir el modelo edad- periodo- cohorte; mostrando la existencia, además, de valores atípicos (*outliers*). En este aspecto, si una tasa de mortalidad por suicidio posee la forma de un círculo negro, ésta se encontrará centrada respecto al resto de respuestas; formando parte, pues, del 80% central de la distribución. Por otra parte, los triángulos hacen referencia a las tasas de mortalidad por suicidio que se encuentran situadas en las colas de la distribución. En este caso, teniendo en cuenta los datos analizados, la mayor parte de las ratios de mortalidad por suicidio se encuentran situadas en la parte central de la distribución. Sin embargo, en los periodos situados entre los años 1985- 1989 y 1995- 1999, las tasas de mortalidad por suicidio de los grupos de población más jóvenes se encuentran situadas en las colas de la distribución. Por otro lado, un número importante de ratios de mortalidad por suicidio pertenecientes al primer periodo (1985-

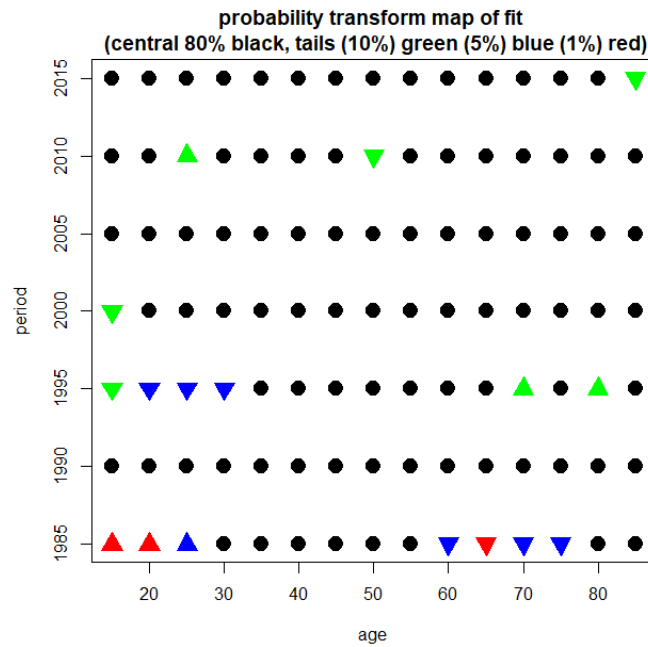


Figura 5.17: Mapa de ajuste de probabilidad.

1989), se encuentran situados en las colas de la distribución; por lo que, existen importantes divergencias, en cuanto a las tasas de mortalidad por suicidio, entre el periodo más antiguo y el resto de periodos que conforman la serie.

Por otra parte, la estimación del modelo edad- periodo- cohorte utilizando la parametrización canónica y el método *detrended*, se muestran en la figura 5.18. De tal forma, dicha gráfica se obtiene mediante la implementación del siguiente comando:

```
> apc.plot.fit(modelo_tasa_suicidios_totales, type = "detrend")
```

De tal forma, las gráficas (a), (b) y (c) muestran las estimaciones de las dobles diferencias en relación con las variables edad, periodo y cohorte. En este caso, no se aprecia ruido alguno en dichas figuras; por lo que, desde un punto de vista gráfico, no sería necesario realizar un análisis de submuestra. Por otro lado, en las figuras (d), (e) y (f) se muestran el nivel y las pendientes asociadas al plano lineal del modelo, cuyos valores fueron introducidos anteriormente. Finalmente, las figuras (g), (h) e (i) muestran las partes no lineales asociadas a los efectos temporales del modelo. Estas se obtienen mediante el cálculo de las dobles sumas de las dobles diferencias empleando el método *detrended*; por lo que, los valores iniciales y finales de cada figura, se encuentran fijados a cero. De tal forma, la gráfica del regresor edad posee una forma cóncava; mientras que, las partes no lineales asociadas al regresor periodo, muestran un notable aumento, en las tasas de mortalidad por suicidio, a partir del periodo 2005- 2009. Finalmente, los elementos no lineales referentes al regresor cohorte muestran un progresivo incremento, en las ratios de mortalidad por suicidio, en relación con las cohortes nacidas a partir de 1940. De tal forma, entre los periodos 1995- 2019, las tasas de mortalidad por suicidio más elevadas se concentraban en las personas con edades comprendidas entre los 40 y los 49 años; coincidiendo, por tanto, con las cohortes nacidas a partir de 1940.

Finalmente, en la figura 5.19 se muestran los residuos obtenidos a partir de la estimación del modelo edad - periodo- cohorte aplicado a las tasas de mortalidad por suicidio. Para obtener esta gráfica, se debe implementar el siguiente comando:

```
> apc.plot.fit.residuals(modelo_tasa_suicidios_totales)
```

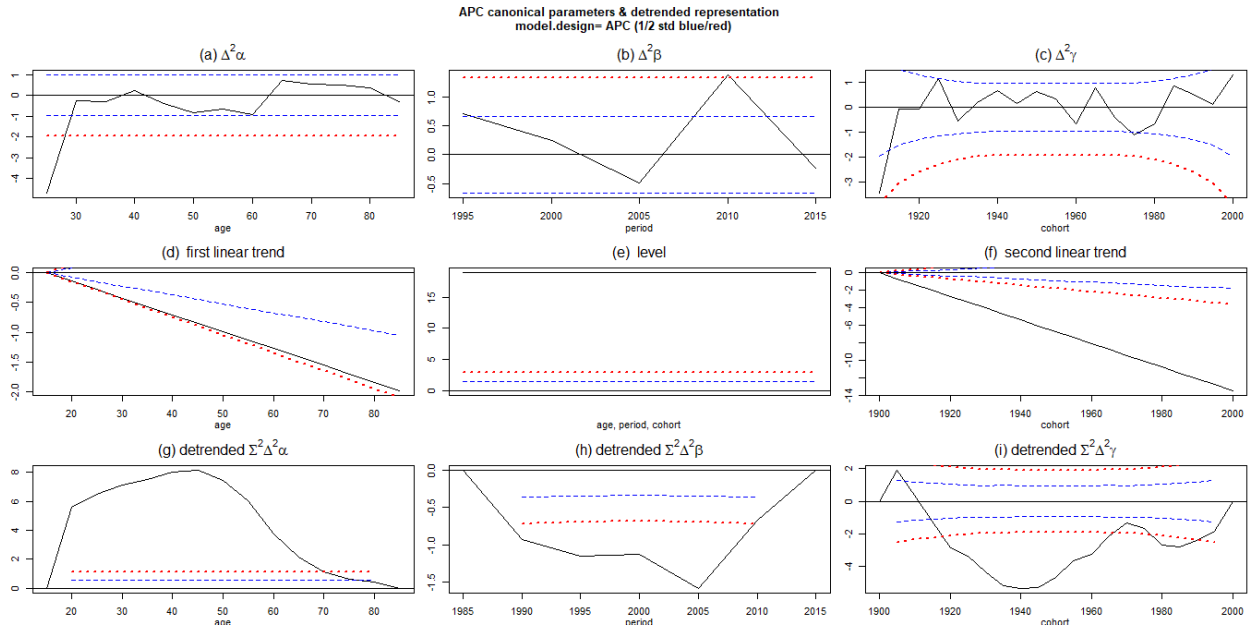



Figura 5.18: Representación de las estimaciones de los parámetros haciendo uso de la metodología *detrended*.

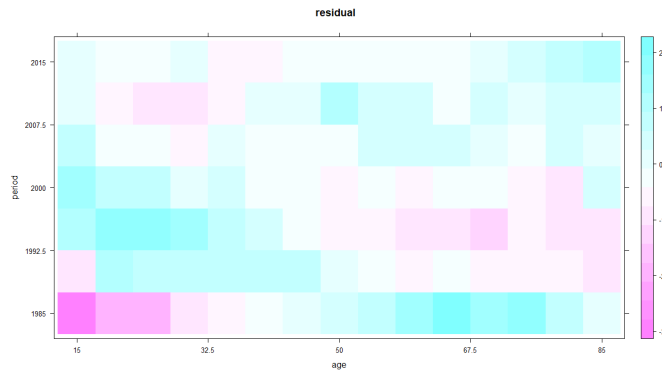


Figura 5.19: Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio.

De tal forma, el modelo presenta un buen ajuste; ya que, la mayoría de los residuos, están representados en colores claros (siendo, estos, los que tienen valores cercanos a cero, en base a la leyenda que aparece en la parte derecha de la figura). Por otro lado, el modelo posee algunos problemas, de carácter leve, relacionados con el ajuste de los grupos de edad situados en los extremos de la distribución (15- 19 años y 85- 89 años). Además, el primer periodo de análisis (1985- 1989) posee residuos elevados, en comparación con el resto de periodos. A pesar de estas cuestiones, el ajuste realizado a través de este modelo presenta un buen comportamiento.

5.2.4. Inferencia sobre el modelo

Tras analizar los resultados obtenidos a partir de la implementación del modelo edad- periodo- cohorte aplicado a las tasas de mortalidad por suicidio, se realizará la correspondiente inferencia sobre el modelo; utilizando, para ello, la *tabla de deviance*. De tal forma, el objetivo de este epígrafe consiste en estudiar si, alguno de los submodelos anidados permite explicar, en mayor medida que el modelo general, las ratios de mortalidad por suicidio de Inglaterra y Gales.

Para obtener la *tabla de deviance*, se debe implementar el siguiente comando:

```
> deviance_table_apcmodel_total_rates=apc.fit.table(modelo_tasa_suicidios_totales,
"gaussian.response")
> deviance_table_apcmodel_total_rates
##      -2logL df.residual LR vs.APC df vs.APC prob(>chi_sq) F vs.APC prob(>F)
## APC 255.349      65      NaN      NaN      NaN      NaN      NaN
## AP 389.181      84 133.832      19      0      8.817      0
## AC 292.461      70  37.113      5      0      5.512      0
## PC 459.889      78 204.540      13      0      30.073      0
## Ad 409.250      89 153.902      24      0      9.021      0
## Pd 512.257      97 256.909      32      0      21.431      0
## Cd 469.937      83 214.588      18      0      24.263      0
## A 471.398      90 216.050      25      0      17.751      0
## P 513.922      98 258.574      33      0      21.145      0
## C 481.272      84 225.924      19      0      25.997      0
## t 518.892     102 263.544      37      0      19.859      0
## tA 545.155     103 289.806      38      0      25.317      0
## tP 520.456     103 265.108      38      0      19.652      0
## tC 536.882     103 281.534      38      0      23.270      0
## 1 546.375     104 291.026      39      0      24.976      0
##      aic
## APC 337.349
## AP 433.181
## AC 364.461
## PC 515.889
## Ad 443.250
## Pd 530.257
## Cd 515.937
## A 503.398
## P 529.922
## C 525.272
## t 526.892
## tA 551.155
## tP 526.456
## tC 542.882
## 1 550.375
```

En este caso, la comparación entre el modelo edad- periodo- cohorte y el resto de submodelos anidados debe realizarse en base a la primera y la última columna de la *tabla de deviance* que se acaba de introducir. De tal forma, la primera columna se corresponde con el valor de menos dos veces el logaritmo de la función de verosimilitud (L); mientras que, la última columna, contiene las cifras correspondientes al criterio de información de Akaike (AIC).

Por tanto, el modelo edad- periodo- cohorte (APC) posee las cifras más bajas tanto en el valor de $-2\log(L)$ como, en el criterio de información de Akaike (AIC). De tal forma, el modelo edad- periodo- cohorte genera el mejor ajuste sobre los datos; siendo las variables edad, periodo y cohorte, por tanto, significativas. Además, como se observaba en la figura 5.17, que representaba las observaciones de las transformadas de probabilidad, el modelo edad- periodo- cohorte se ajustaba, de forma correcta, a los datos; ya que, la mayoría de las cifras, se concentraban en la parte central de la distribución.

Finalmente, es conveniente comparar los resultados obtenidos al aplicar el modelo Poisson (implementado a las cifras brutas de suicidios) y el modelo gaussiano (aplicado a las tasas de mortalidad por suicidio, por 100000 habitantes). De tal forma, si se comparan las gráficas de residuos obtenidas mediante la aplicación del modelo Poisson (figura 5.8) y el modelo gaussiano (figura 5.19) se observa que, el modelo gaussiano aplicado a las ratios de mortalidad por suicidio, genera menores cifras de residuos que el modelo Poisson. Además,

si se comparan los gráficos de ajuste de probabilidad de ambos modelos (figura 5.7 en el caso del modelo Poisson y gráfico 5.17 en relación con el modelo gaussiano); se observa que, la mayor parte de los datos sobre los que se ajusta el modelo gaussiano, se encuentran situados en la parte del 80% central de la distribución, a diferencia de lo que ocurre en el caso del modelo Poisson. Por tanto, el modelo gaussiano presenta una mayor capacidad de ajuste, en comparación con el modelo Poisson; por lo que, se usará este modelo a la hora de comparar las tasas de mortalidad por suicidio entre hombres y mujeres.

5.3. Aplicación de un *AC model* a las tasas de mortalidad por suicidio

En el apartado anterior se obtuvo la *tabla de deviance* aplicada a la familia de modelos gaussianos; obteniéndose que, el modelo edad- periodo- cohorte que ajustaba las tasas de mortalidad por suicidio, poseía los menores valores referentes tanto al estadístico $-2\log(L)$ (menos dos veces el logaritmo de la función de verosimilitud), como al criterio de información de Akaike (AIC). Sin embargo, el submodelo edad- cohorte (columna AC) poseía valores ligeramente superiores, en cuanto a dichos indicadores, respecto al modelo general (APC), circunstancia que hacía preferible a este último modelo. De hecho, el valor de $-2\log(L)$ era igual a 255,349 para el modelo APC; mientras que, dicho indicador era igual a 292,461 para el modelo edad-cohorte. Por otro lado, el criterio de información de Akaike tomaba valores iguales a 337,349 y 364,4614 para los modelos edad- periodo- cohorte (APC) y edad- cohorte (AC), respectivamente. Por tanto, las diferencias existentes entre ambos modelos, al menos en cuanto a estos dos indicadores, no parecen ser importantes; por lo que, en base al principio de parsimonia; se investigará si, la variable periodo, posee una destacada relevancia a la hora de ajustar el modelo. Para tal fin, se aplicará el modelo edad- cohorte a las tasas de mortalidad por suicidio.

Los comandos ejecutados para obtener dicho modelo se muestran a continuación:

```
> modelo_tasa_suicidios_totales_ACmodel=apc.fit.model(tasa_suicidios_totales,
model.family = "gaussian.response",model.design = "AC")
```

De tal forma, los coeficientes canónicos obtenidos a través de la implementación de dicho modelo son los siguientes:

```
> modelo_tasa_suicidios_totales_ACmodel$coefficients.canonical
##              Estimate Std. Error   t value    Pr(>|t|)
## level          14.69389668  0.6598275 22.26929865 7.334009e-110
## age slope       -1.51208535  0.6484067 -2.33200128 1.970062e-02
## cohort slope    -0.83925128  0.6485434 -1.29405575 1.956462e-01
## DD_age_25       -4.75033291  1.1291131 -4.20713635 2.586271e-05
## DD_age_30       -0.31455800  1.1197477 -0.28091864 7.787728e-01
## DD_age_35       -0.34328304  1.1162881 -0.30752189 7.584462e-01
## DD_age_40        0.19436582  1.1144587  0.17440379 8.615481e-01
## DD_age_45       -0.40991700  1.1133219 -0.36819271 7.127295e-01
## DD_age_50       -0.81368603  1.1125466 -0.73137253 4.645516e-01
## DD_age_55       -0.67632020  1.1123111 -0.60803153 5.431666e-01
## DD_age_60       -0.92107442  1.1125466 -0.82789738 4.077286e-01
## DD_age_65        0.72255042  1.1133219  0.64900405 5.163358e-01
## DD_age_70        0.51634438  1.1144587  0.46331406 6.431393e-01
## DD_age_75        0.48940585  1.1162881  0.43842251 6.610800e-01
## DD_age_80        0.34084391  1.1197477  0.30439350 7.608281e-01
## DD_age_85       -0.36350097  1.1291131 -0.32193494 7.475020e-01
## DD_cohort_1910 -3.17216634  2.2103783 -1.43512374 1.512518e-01
## DD_cohort_1915  0.06174744  1.7371320  0.03554562 9.716447e-01
## DD_cohort_1920 -0.09318220  1.4866660 -0.06267864 9.500224e-01
## DD_cohort_1925  1.38735151  1.3224886  1.04904613 2.941569e-01
```

```
## DD_cohort_1930 -0.43902063  1.2036113 -0.36475283  7.152959e-01
## DD_cohort_1935  0.02666975  1.1258146  0.02368929  9.811004e-01
## DD_cohort_1940  0.69310729  1.1125462  0.62299190  5.332898e-01
## DD_cohort_1945  0.17741944  1.1123978  0.15949281  8.732806e-01
## DD_cohort_1950  0.66181386  1.1123556  0.59496610  5.518661e-01
## DD_cohort_1955  0.35824815  1.1123449  0.32206570  7.474029e-01
## DD_cohort_1960 -0.65609079  1.1123556 -0.58982110  5.553106e-01
## DD_cohort_1965  0.80635327  1.1123978  0.72487854  4.685265e-01
## DD_cohort_1970 -0.38660980  1.1125462 -0.34749999  7.282157e-01
## DD_cohort_1975 -1.24533188  1.1258146 -1.10616074  2.686570e-01
## DD_cohort_1980 -0.52896379  1.2036113 -0.43948058  6.603134e-01
## DD_cohort_1985  0.93183412  1.3224886  0.70460656  4.810551e-01
## DD_cohort_1990  0.58179017  1.4866660  0.39133886  6.955468e-01
## DD_cohort_1995  0.48283391  1.7371320  0.27794888  7.810516e-01
## DD_cohort_2000  1.23677764  2.2103783  0.55953212  5.757986e-01
```

Finalmente, con el objetivo de comparar ambos modelos, se introducirán los gráficos de transformadas de probabilidad y de residuos concernientes al modelo edad- cohorte.

En este aspecto, en la figura 5.20, se muestra el gráfico de transformadas de probabilidad referente al modelo edad- cohorte. Dicha figura se ha obtenido a partir de la implementación del siguiente comando:

```
> apc.plot.fit.pt(modelo_tasa_suicidios_totales_ACmodel)
```

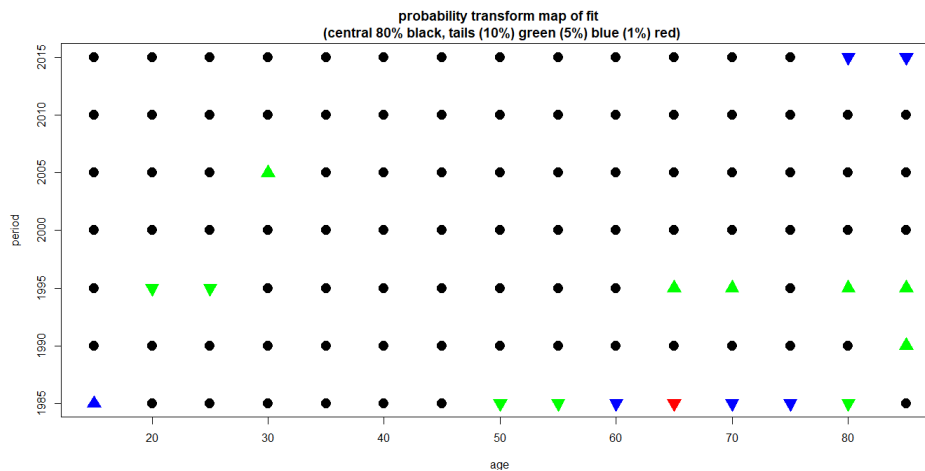


Figura 5.20: Transformadas de probabilidad obtenidas al aplicar un modelo edad- cohorte a las tasas de mortalidad por suicidio.

Si se compara la figura 5.20 (modelo edad- cohorte) con la gráfica 5.17 (modelo edad- periodo- cohorte) se observa que, en ambos ajustes, las transformadas de probabilidad se sitúan, mayoritariamente, en el 80 % central de la distribución. Sin embargo, el ajuste edad- cohorte genera una menor cantidad de datos en las colas; ya que, el número de triángulos azules y rojos contenidos en la figura 5.20 (modelo edad- cohorte) es menor que la cifra de triángulos azules y rojos presentes en la gráfica 5.17 (modelo edad- periodo- cohorte). Por tanto, en base a esta gráfica, el modelo edad- cohorte podría generar mejores resultados que el modelo edad- periodo- cohorte (sin embargo, se deberían implementar análisis estadísticos que corroborasen esta última idea).

Para continuar con el análisis, se introducirá, en la figura 5.21, la representación de los residuos asociados al modelo edad- cohorte. Para ello, se implementará el siguiente comando:

```
> apc.plot.fit.residuals(modelo_tasa_suicidios_totales_ACmodel)
```

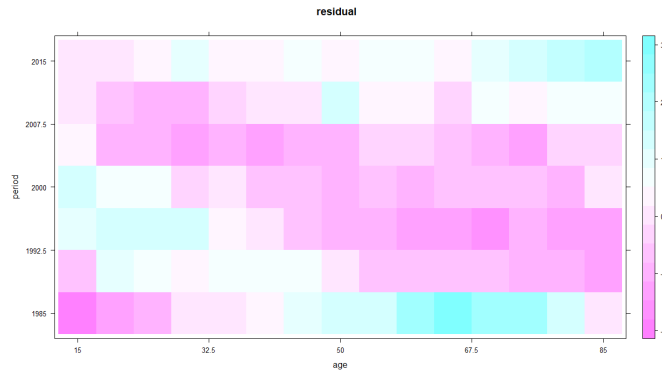


Figura 5.21: Representación gráfica de los residuos obtenidos al implementar un modelo edad- cohorte a las tasas de mortalidad por suicidio.

Si se comparan los residuos de la figura 5.21 (modelo edad- cohorte) con los obtenidos a partir de la estimación del modelo general (figura 5.19); se observa que, los residuos obtenidos mediante el ajuste edad- cohorte, son notablemente superiores a los arrojados por la estimación del modelo edad- periodo- cohorte. Por tanto, el ajuste realizado a través del modelo general reviste de una mayor capacidad de ajuste, en comparación con el submodelo edad- cohorte; por lo que, el regresor periodo, resulta ser significativo a la hora de ajustar las tasas de mortalidad por suicidio. De tal forma, dicha conclusión también fue obtenida, anteriormente, a través del estudio de la *tabla de deviance*.

5.4. Estudio comparativo de las tasas de mortalidad por suicidio de hombres y mujeres

Las diferencias existentes en las cifras de suicidios de varones y mujeres son notables, como bien se señaló en el epígrafe 2.3 (ver tablas 2.1 y 2.2); por lo que, en este apartado, se analizarán dichas divergencias. Para ello, se utilizará el modelo gaussiano; ya que, este último modelo arrojó mejores resultados que los que se obtuvieron mediante el ajuste de Poisson. De tal forma, al emplear el modelo gaussiano, se utilizarán las tasas de mortalidad por suicidio; que se encuentran contenidas, a su vez, en la tabla 2.2.

Los pasos que se realizarán en este apartado son idénticos a los que se implementaron en el caso del análisis de las tasas de mortalidad por suicidio agregadas; aunque, en este caso, la interpretación se centrará en estudiar las divergencias existentes, entre ambos sexos, en relación con las tasas de mortalidad por suicidio.

5.4.1. Creación de los objetos de datos

En primer lugar, se deben introducir los datos que se utilizarán en el estudio en el software R. Para ello, se ha utilizado la función *read_excel*; que se encuentra disponible, a su vez, en la librería *readxl*; ya que, los datos del estudio, se encontraban contenidos en un archivo *.xls*. Una vez introducidos los datos, estos han sido depurados y transformados en matrices; siguiendo, para ello, los pasos contenidos en el anexo de este documento (apéndice A.1.2).

De tal forma, tras la introducción de las bases de datos a analizar, se deben crear los correspondientes objetos de tipología *apc.data.list* (uno para varones y otro para mujeres). Para ello, se deben ejecutar los siguientes comandos:

```
> library (apc) # Se carga la librería
# Se deben transformar los dataframes de las tasas en matrices
> tasa_suicidios_hombres_matrix=as.matrix(t_Tasas_Suicidios_Males)
> tasa_suicidios_mujeres_matrix=as.matrix(t_Tasas_Suicidios_Females)
```

```

# Se crean los objetos correspondientes con la función apc.data.list
> tasa_suicidios_hombres=apc.data.list(tasa_suicidios_hombres_matrix,"AP",age1=15, per1 = 1985,
unit = 5, label="Suicide mortality rates England and Wales (period 1985- 2019) for men")
> tasa_suicidios_hombres
## $response
##      1985-1989 1990-1994 1995-1999 2000-2004 2005-2009 2010-2014 2015-2019
## 15-19      8.00      8.40      8.12      7.24      5.40      5.60      7.34
## 20-24     18.44     22.70     21.56     16.76     13.26     13.30     14.28
## 25-29     18.90     23.22     24.60     20.72     15.42     14.62     15.98
## 30-34     21.10     22.12     25.18     22.76     18.62     16.70     18.06
## 35-39     20.90     23.32     21.84     22.64     21.64     20.22     18.32
## 40-44     21.32     23.00     22.36     21.10     21.54     24.06     22.06
## 45-49     21.36     21.96     20.82     20.60     19.36     23.12     23.88
## 50-54     20.20     20.14     17.76     18.12     18.50     22.28     21.28
## 55-59     21.36     17.82     16.30     15.94     16.68     19.12     19.06
## 60-64     20.62     15.60     14.12     12.44     12.96     14.92     15.40
## 65-69     20.66     16.12     12.68     12.14     10.98     10.86     11.72
## 70-74     21.88     15.58     12.66     12.70     10.72     11.52     10.84
## 75-79     26.46     18.88     15.96     14.18     11.68     11.64     11.54
## 80-84     35.16     26.22     20.50     16.58     16.20     14.46     14.62
## 85-89     33.28     34.74     27.62     24.10     19.50     18.16     18.08
##
## $dose
## NULL
##
## $data.format
## [1] "AP"
##
## $age1
## [1] 15
##
## $per1
## [1] 1985
##
## $coh1
## [1] 5
##
## $unit
## [1] 5
##
## $per.zero
## NULL
##
## $per.max
## NULL
##
## $time.adjust
## [1] 0
##
## $label
## [1] "Suicide mortality rates England and Wales (period 1985- 2019) for men"
##
## $n.decimal
## NULL
> tasa_suicidios_mujeres=apc.data.list(tasa_suicidios_mujeres_matrix,"AP",age1=15, per1 = 1985,

```

```

unit = 5, label="Suicide mortality rates England and Wales (period 1985- 2019) for women")
> tasa_suicidios_mujeres
## $response
##      1985-1989 1990-1994 1995-1999 2000-2004 2005-2009 2010-2014 2015-2019
## 15-19      2.62      2.26      2.76      2.40      2.02      1.98      3.40
## 20-24      4.32      4.58      4.60      4.32      3.02      3.52      4.44
## 25-29      5.04      5.06      5.38      5.42      4.36      3.70      4.74
## 30-34      6.30      5.88      5.40      5.30      4.76      4.42      5.08
## 35-39      6.78      5.92      6.32      6.08      5.28      5.58      5.28
## 40-44      7.54      6.62      6.94      6.18      5.40      5.82      5.98
## 45-49      9.84      7.42      6.34      6.78      6.68      6.62      6.92
## 50-54     11.68      7.58      6.44      6.76      6.66      6.86      7.28
## 55-59     10.98      7.48      6.44      6.44      6.02      5.94      5.86
## 60-64     11.84      7.62      5.16      5.32      5.24      4.62      5.38
## 65-69     12.30      7.68      5.96      5.06      4.50      4.08      4.40
## 70-74     12.96      8.40      6.66      5.34      3.70      3.96      3.98
## 75-79     13.50      9.02      6.76      6.14      4.14      3.94      4.00
## 80-84     12.30      9.38      7.48      6.02      5.44      4.86      3.64
## 85-89     11.42      9.54      8.20      7.96      5.02      5.20      4.42
##
## $dose
## NULL
##
## $data.format
## [1] "AP"
##
## $age1
## [1] 15
##
## $per1
## [1] 1985
##
## $coh1
## [1] 5
##
## $unit
## [1] 5
##
## $per.zero
## NULL
##
## $per.max
## NULL
##
## $time.adjust
## [1] 0
##
## $label
## [1] "Suicide mortality rates England and Wales (period 1985- 2019) for women"
##
## $n.decimal
## NULL

```

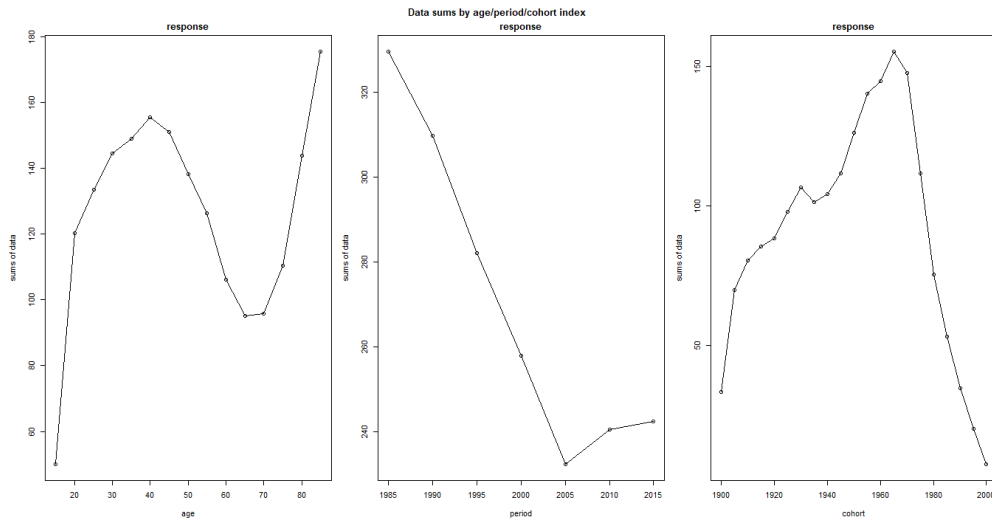
Por tanto, se han seguido los mismos pasos que fueron ejecutados en los epígrafes anteriores.

5.4.2. Exploración de los datos

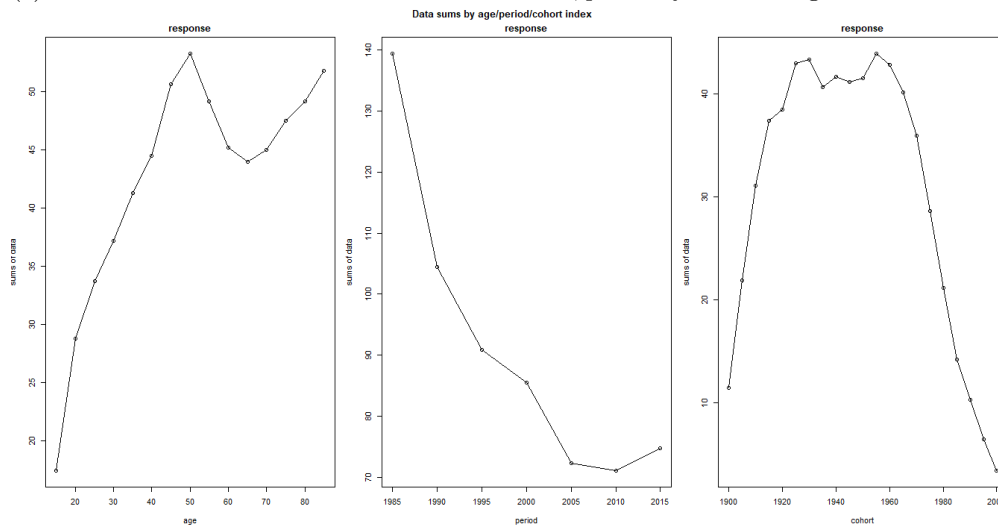
Una vez introducidos los objetos de datos mediante la función *apc.data.list*, se procederá con el análisis gráfico de los mismos. Para ello, se ejecutarán las siguientes instrucciones:

```
> apc.plot.data.all(tasa_suicidios_hombres)
> apc.plot.data.all(tasa_suicidios_mujeres)
```

Los resultados gráficos obtenidos se irán mostrando progresivamente, junto a la interpretación de los mismos.



(a) Suma de los datos en función de las variables edad, periodo y cohorte. Segmento de varones.



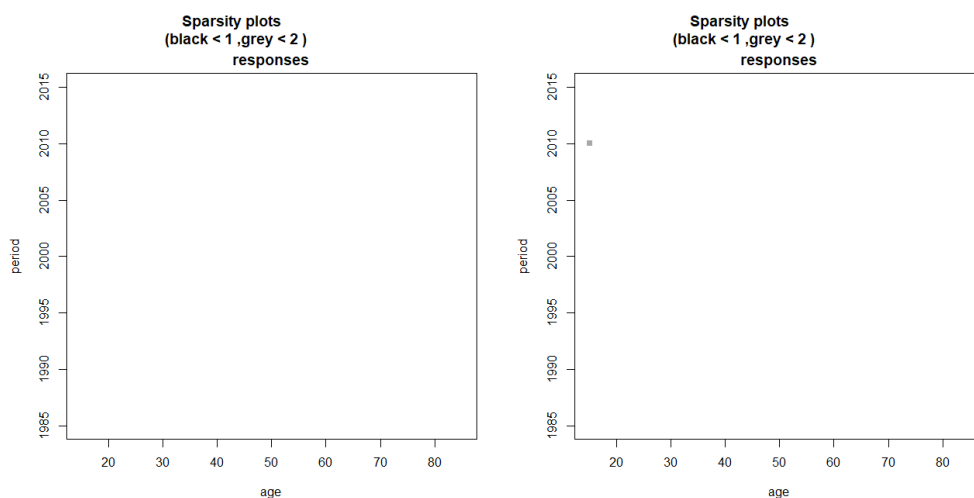
(b) Suma de los datos en función de las variables edad, periodo y cohorte. Segmento de mujeres.

Figura 5.22: Suma de los datos en función de las variables edad, periodo y cohorte; segmentando, a su vez, por sexos.

En primer lugar, en la figura 5.22 se muestran las sumas de los datos en función de los regresores edad, periodo y cohorte; segmentando, a su vez, en función del sexo del fallecido. De tal forma, ambas figuras presentan importantes similitudes; aunque, existen algunas divergencias de notable interés. En este aspecto, en relación con el regresor edad, las tasas de mortalidad por suicidio de varones y mujeres crecen, respectivamente, entre los 15- 49 años y 15- 54 años. A partir de dichos rangos de edad, las ratios de mortalidad por

suicidio comienzan a decrecer, con independencia del sexo; volviendo a incrementarse, en ambas gráficas, a partir del grupo con edades comprendidas entre los 70- 74 años. Por otro lado, en relación con la variable periodo, se observa una reducción destacada de las tasas de mortalidad por suicidio entre los periodos 1985- 2009, en el caso de los varones, y 1985- 2015, en el caso del segmento de mujeres. De tal forma, durante los últimos años de la serie, se ha producido un aumento de las tasas de mortalidad por suicidio, con independencia del sexo. Finalmente, las cohortes nacidas entre 1940 y 1980 presentan las mayores tasas de mortalidad por suicidio en el segmento de varones; mientras que, en el caso de las mujeres, las cohortes de nacimiento con mayores ratios de mortalidad por suicidio se corresponden con aquellas nacidas entre 1920 y 1980.

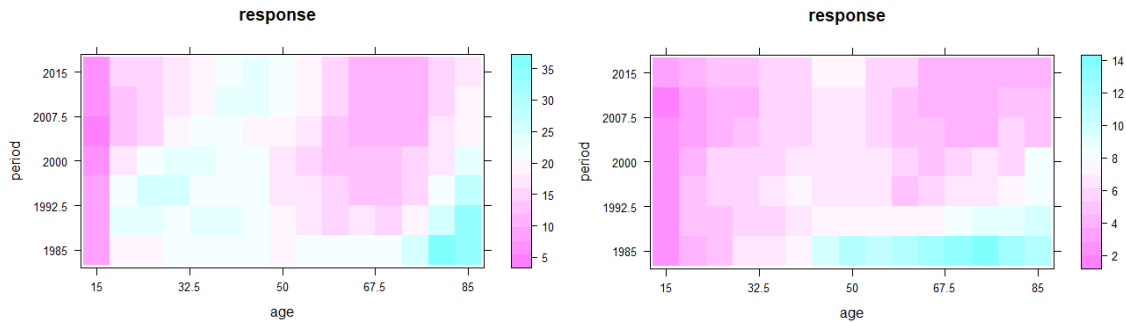
Por otro lado, en las figuras 5.23a y 5.23b se muestran las gráficas tipo *sparsity* para varones y mujeres. En el caso de las personas de sexo masculino, la gráfica se encuentra vacía; ya que, no existen tasas de mortalidad por suicidio con valores inferiores al 2%. Sin embargo, en el caso de las mujeres, aparece un punto gris en el grupo con edades comprendidas entre los 15- 19 años (periodo 2010- 2014); ya que, en este caso, las tasas de mortalidad por suicidio, para dicho grupo de edad, tomaban un valor del 1,98 %; (inferior, pues, al 2%). A pesar de este hecho, no existen problemas graves de dispersión en los datos.



(a) Gráfico tipo *sparsity* para las tasas de mortalidad por suicidio de varones. (b) Gráfico tipo *sparsity* para las tasas de mortalidad por suicidio de mujeres.

Figura 5.23: Gráficos tipo *sparsity* para las tasas de mortalidad por suicidio de varones y mujeres.

Por otro lado, en las figuras 5.24a y 5.24b se muestra la distribución de las tasas de mortalidad por suicidio de varones y mujeres. De tal forma, si se analiza la leyenda de ambas figuras se observa que, las tasas de mortalidad por suicidio de varones, son superiores a las ratios de mortalidad por suicidio de mujeres. Además, en ambos casos, las personas con edades comprendidas entre los 15 y 19 años, poseen las tasas de mortalidad por suicidio más reducidas. Por otro lado, las ratios de mortalidad por suicidio más elevadas, considerando el segmento de los varones, se concentran en los grupos con edades comprendidas entre los 35- 49 años y los 80- 89 años. Por otra parte, en el caso de las mujeres, las tasas de mortalidad por suicidio con las cifras más elevadas, pertenecen a aquellas personas con edades comprendidas entre los 40- 54 años y los 85- 89 años.



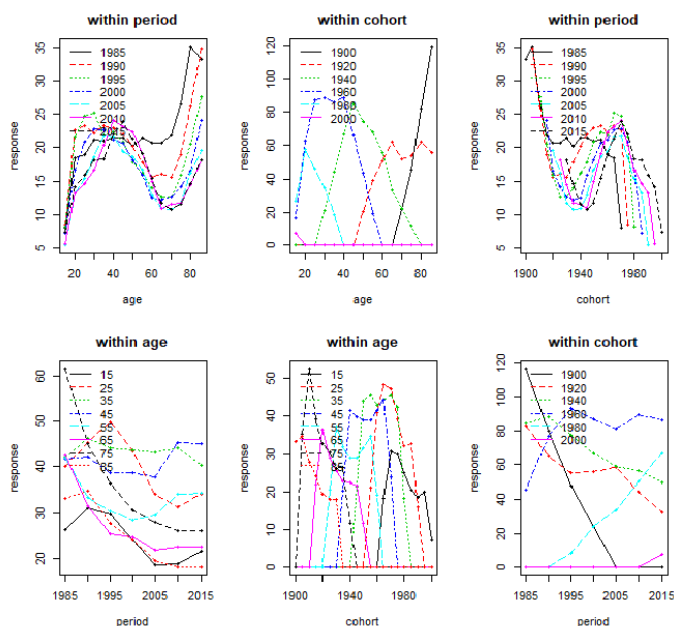
(a) Distribución de las tasas de mortalidad por suicidio de varones, en función de las variables edad y periodo. (b) Distribución de las tasas de mortalidad por suicidio de mujeres, en función de las variables edad y periodo.

Figura 5.24: Distribución de las tasas de mortalidad por suicidio en función de las variables edad y periodo.

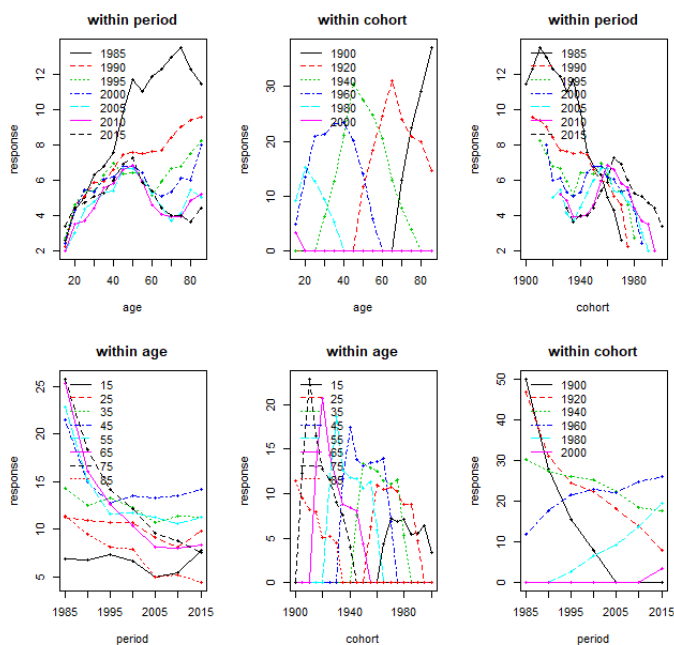
Finalmente, en las figuras 5.25a y 5.25b se muestran los gráficos de las tasas de mortalidad por suicidio de varones y mujeres; empleando, para ello, dos índices temporales. Se procederá a describir, de forma individual y comparada, cada uno de dichos gráficos:

- **Primer gráfico.** Esta figura representa las tasas de mortalidad por suicidio en función de las variables edad y periodo. En este aspecto, independientemente del sexo, se observa un aumento de las ratios de mortalidad por suicidio en la vejez. Además, las tasas de mortalidad por suicidio se han reducido a lo largo del periodo de análisis, tanto en el caso de los varones como en el de las mujeres.
- **Segundo gráfico.** En este caso, se representan las ratios de mortalidad por suicidio en función de los regresores edad y cohorte. De tal forma, se observan resultados similares en hombres y mujeres; ya que, las cohortes más antiguas, concentran las tasas de mortalidad por suicidio en edades avanzadas; mientras que, las cohortes de nacimiento más recientes, agrupan las mayores ratios de mortalidad por suicidio en los grupos de población más jóvenes.
- **Tercer gráfico.** Esta gráfica refleja las tasas de mortalidad por suicidio considerando las variables cohorte y periodo; poseyendo resultados similares, en este aspecto, en los segmentos de varones y mujeres. De tal forma, las cohortes más antiguas, concentran las tasas de mortalidad por suicidio en los primeros periodos de análisis; mientras que, las cohortes de nacimiento más recientes, agrupan las tasas de mortalidad por suicidio en los últimos periodos de la serie.
- **Cuarta gráfica.** En este caso, se representan las ratios de mortalidad por suicidio en función de las variables edad y periodo. En cuanto a la tendencia, los resultados obtenidos son similares, con independencia del sexo; aunque, las tasas de mortalidad por suicidio de varones, son notablemente superiores a las ratios de mortalidad por suicidio de mujeres. En el caso de los varones, las tasas de mortalidad por suicidio de las personas con edades comprendidas entre los 15- 34 años y 65- 89 años, se han visto reducidas a lo largo del periodo de análisis; mientras que, las ratios de mortalidad por suicidio de los individuos con edades comprendidas entre los 35 y 64 años, se han visto incrementadas. Por otro lado, en el caso de las mujeres, las ratios de mortalidad por suicidio se han reducidas, de forma notable, durante el periodo de análisis, con la excepción del grupo de personas con edades comprendidas entre los 15 y 24 años.
- **Quinta gráfica.** Estas figuras reflejan las tasas de mortalidad por suicidio en función de las variables edad y cohorte, poseyendo notables similitudes con independencia del sexo. Además, se observa una reducción de las ratios de mortalidad por suicidio a medida que las cohortes de nacimiento alcanzan periodos más recientes.
- **Sexta gráfica.** En este caso, se representan las tasas de mortalidad por suicidio en función de los regresores periodo y cohorte, obteniéndose resultados similares con independencia del sexo. En este aspecto, las cohortes de nacimiento más antiguas han reducido su importancia, en relación con las tasas

de mortalidad por suicidio, a medida que aumenta el periodo. Por otro lado, las cohortes de nacimiento más recientes han adquirido, progresivamente, una mayor importancia en las tasas de mortalidad por suicidio; estando motivado, este hecho, por cuestiones demográficas.



(a) Gráfico de las tasas de mortalidad por suicidio de varones utilizando dos índices temporales.



(b) Gráfico de las tasas de mortalidad por suicidio de mujeres utilizando dos índices temporales.

Figura 5.25: Gráficos de las tasas de mortalidad por suicidio utilizando dos índices temporales.

5.4.3. Estimación de los modelos

En primer lugar, se estimarán los modelos edad- periodo- cohorte aplicados a las tasas de mortalidad por suicidio de hombres y mujeres; considerando, para ello, la familia de modelos gaussianos.

La estimación de estos modelos se obtendrá mediante la implementación de los siguientes comandos:

```
> modelo_tasa_suicidios_males=apc.fit.model(tasa_suicidios_hombres,
model.family = "gaussian.response",model.design = "APC")
> modelo_tasa_suicidios_males$coefficients.canonical
##          Estimate Std. Error   t value   Pr(>|t|)
## level          20.3846217  0.8434216 24.1689578 4.719256e-129
## age slope       -1.8407957  0.9445270 -1.9489074 5.130648e-02
## cohort slope    -0.8862303  0.9451299 -0.9376810 3.484084e-01
## DD_age_25       -8.2099775  1.3612479 -6.0312143 1.627322e-09
## DD_age_30       -0.3021956  1.3495449 -0.2239241 8.228164e-01
## DD_age_35       -0.7680771  1.3452691 -0.5709468 5.680357e-01
## DD_age_40        0.5462593  1.3430214  0.4067391 6.841996e-01
## DD_age_45       -1.2197210  1.3416302 -0.9091335 3.632797e-01
## DD_age_50       -1.0265227  1.3406843 -0.7656707 4.438723e-01
## DD_age_55       -0.1361204  1.3403880 -0.1015530 9.191115e-01
## DD_age_60       -1.6486884  1.3406843 -1.2297364 2.187958e-01
## DD_age_65        1.1141744  1.3416302  0.8304631 4.062770e-01
## DD_age_70        1.0667832  1.3430214  0.7943158 4.270116e-01
## DD_age_75        1.1789539  1.3452691  0.8763703 3.808288e-01
## DD_age_80        1.7543243  1.3495449  1.2999377 1.936223e-01
## DD_age_85        0.2628625  1.3612479  0.1931041 8.468775e-01
## DD_period_1995 -0.1566015  0.9193038 -0.1703480 8.647365e-01
## DD_period_2000 -0.1676433  0.9159995 -0.1830168 8.547848e-01
## DD_period_2005 -0.3820084  0.9153522 -0.4173348 6.764335e-01
## DD_period_2010  2.0299628  0.9159995  2.2161177 2.668344e-02
## DD_period_2015 -0.6513278  0.9193038 -0.7085012 4.786341e-01
## DD_cohort_1910 -9.8887536  2.6825431 -3.6863354 2.275064e-04
## DD_cohort_1915  2.3511413  2.1021260  1.1184588 2.633711e-01
## DD_cohort_1920  1.4210674  1.7969212  0.7908346 4.290405e-01
## DD_cohort_1925  1.6744094  1.5974530  1.0481744 2.945582e-01
## DD_cohort_1930 -0.2207441  1.4532900 -0.1518927 8.792716e-01
## DD_cohort_1935  0.4874550  1.3576561  0.3590416 7.195640e-01
## DD_cohort_1940  0.6753872  1.3406879  0.5037617 6.144288e-01
## DD_cohort_1945  0.2910137  1.3404945  0.2170943 8.281349e-01
## DD_cohort_1950  1.1187344  1.3404405  0.8346021 4.039418e-01
## DD_cohort_1955  0.1462622  1.3404271  0.1091161 9.131104e-01
## DD_cohort_1960 -0.9773292  1.3404405 -0.7291104 4.659341e-01
## DD_cohort_1965  1.5792951  1.3404945  1.1781437 2.387393e-01
## DD_cohort_1970 -1.0605207  1.3406879 -0.7910273 4.289281e-01
## DD_cohort_1975 -2.3917969  1.3576561 -1.7617105 7.811823e-02
## DD_cohort_1980 -0.7765981  1.4532900 -0.5343725 5.930839e-01
## DD_cohort_1985  1.8768726  1.5974530  1.1749157 2.400284e-01
## DD_cohort_1990  0.3866172  1.7969212  0.2151554 8.296462e-01
## DD_cohort_1995  0.4215003  2.1021260  0.2005115 8.410806e-01
## DD_cohort_2000  1.9737685  2.6825431  0.7357826 4.618630e-01

> modelo_tasa_suicidios_females=apc.fit.model(tasa_suicidios_mujeres,
model.family = "gaussian.response",model.design = "APC")
> modelo_tasa_suicidios_females$coefficients.canonical
##          Estimate Std. Error   t value   Pr(>|t|)
```

```

## level          10.38034945  0.4045519 25.65888325 3.365240e-145
## age slope      -2.75279025  0.4530476 -6.07616069 1.230941e-09
## cohort slope  -2.40000514  0.4533368 -5.29408821 1.196117e-07
## DD_age_25      -1.05399891  0.6529301 -1.61425990 1.064711e-01
## DD_age_30      -0.24799754  0.6473167 -0.38311621 7.016336e-01
## DD_age_35       0.07983087  0.6452658  0.12371781 9.015387e-01
## DD_age_40      -0.11355919  0.6441877 -0.17628278 8.600718e-01
## DD_age_45       0.35419045  0.6435204  0.55039505 5.820484e-01
## DD_age_50      -0.61982597  0.6430667 -0.96385954 3.351164e-01
## DD_age_55      -1.16150870  0.6429246 -1.80660179 7.082439e-02
## DD_age_60      -0.18062973  0.6430667 -0.28088801 7.787963e-01
## DD_age_65       0.28662898  0.6435204  0.44540775 6.560251e-01
## DD_age_70       0.01943020  0.6441877  0.03016232 9.759376e-01
## DD_age_75       0.07384134  0.6452658  0.11443554 9.088925e-01
## DD_age_80      -0.12135347  0.6473167 -0.18747155 8.512909e-01
## DD_age_85       0.24347210  0.6529301  0.37289152 7.092292e-01
## DD_period_1995 1.44211474  0.4409492  3.27047834 1.073658e-03
## DD_period_2000 0.53810863  0.4393643  1.22474372 2.206718e-01
## DD_period_2005 -0.63395558  0.4390538 -1.44391332 1.487634e-01
## DD_period_2010 0.67224121  0.4393643  1.53003159 1.260089e-01
## DD_period_2015 0.20675594  0.4409492  0.46888837 6.391494e-01
## DD_cohort_1910 -0.62186228  1.2866967 -0.48330136 6.288818e-01
## DD_cohort_1915 -0.68965436  1.0082964 -0.68397978 4.939879e-01
## DD_cohort_1920 -0.63583890  0.8619033 -0.73771490 4.606877e-01
## DD_cohort_1925 1.00680198  0.7662272  1.31397314 1.888552e-01
## DD_cohort_1930 -0.56373957  0.6970786 -0.80871732 4.186778e-01
## DD_cohort_1935 0.17997136  0.6512073  0.27636569 7.822672e-01
## DD_cohort_1940 0.64001704  0.6430684  0.99525500 3.196123e-01
## DD_cohort_1945 0.03191676  0.6429757  0.04963915 9.604099e-01
## DD_cohort_1950 0.30398192  0.6429498  0.47279264 6.363611e-01
## DD_cohort_1955 0.48503914  0.6429433  0.75440419 4.506066e-01
## DD_cohort_1960 -0.30358953  0.6429498 -0.47218235 6.367966e-01
## DD_cohort_1965 0.04707140  0.6429757  0.07320868 9.416401e-01
## DD_cohort_1970 0.16612414  0.6430684  0.25833044 7.961519e-01
## DD_cohort_1975 0.14823799  0.6512073  0.22763564 8.199295e-01
## DD_cohort_1980 -0.44619821  0.6970786 -0.64009738 5.221093e-01
## DD_cohort_1985 -0.16177057  0.7662272 -0.21112611 8.327889e-01
## DD_cohort_1990 0.65130867  0.8619033  0.75566329 4.498511e-01
## DD_cohort_1995 -0.12879067  1.0082964 -0.12773096 8.983619e-01
## DD_cohort_2000 0.68908099  1.2866967  0.53554266 5.922747e-01

```

En este aspecto, las tres primeras filas de las tablas de coeficientes canónicos que se acaban de estimar, posibilitan la definición del plano lineal del modelo. En concreto, dicho plano lineal estaría definido por los tres primeros elementos de la columna *Estimate*; que poseen, a su vez, sus correspondientes niveles de significación (columna $Pr(> |z|)$).

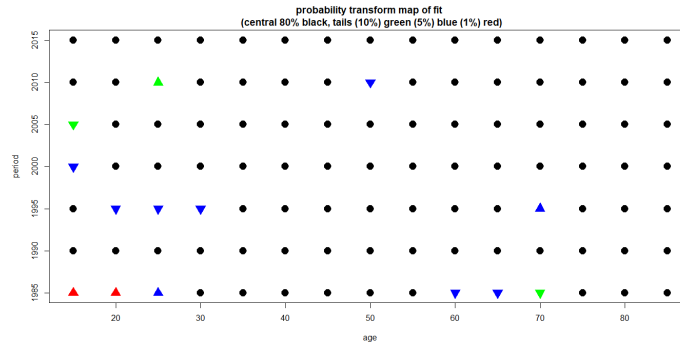
De tal forma, considerando las tasas de mortalidad por suicidio de varones, el nivel sería igual a $\hat{v}_0 = 20,3846217$ ($Pr(> |z|) = 4,719256e - 129$); mientras que, las pendientes de los regresores edad y cohorte serían iguales, respectivamente, a $\hat{v}_{\text{age}} = -1,8407957$ ($Pr(> |z|) = 5,130648e - 02$) y $\hat{v}_{\text{cohort}} = -0,8862303$ ($Pr(> |z|) = 3,484084e - 01$). Por tanto, considerando un nivel de significación estándar del 5%, las pendientes de las variables edad y cohorte no serían significativas.

Por otro lado, considerando las ratios de mortalidad por suicidio de mujeres, el nivel del plano lineal sería igual a $\hat{v}_0 = 10,38034945$ ($Pr(> |z|) = 3,365240e - 145$); mientras que, las pendientes de los regresores edad y cohorte serían iguales, respectivamente, a $\hat{v}_{\text{age}} = -2,75279025$ ($Pr(> |z|) = 1,230941e - 09$) y $\hat{v}_{\text{cohort}} = -2,40000514$ ($Pr(> |z|) = 1,196117e - 07$). Por tanto, considerando un nivel de significación

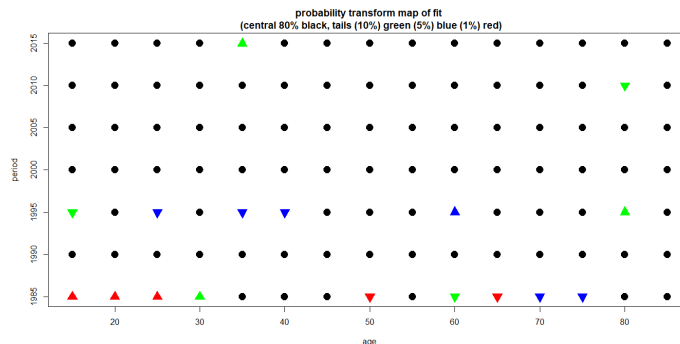
estándar del 5%, las pendientes de las variables edad y cohorte serían significativas, al igual que el nivel.

Tras introducir los coeficientes de correlación canónicos asociados a los modelos edad- periodo- cohorte, se procederá con el análisis gráfico del modelo. Para ello, se realizará una comparación de los resultados obtenidos en función del sexo, siguiendo pasos equivalentes a los implementados a lo largo de este epígrafe. De tal forma, en primer lugar, se introducirán, en la figura 5.26, las gráficas de transformadas de probabilidad. Estas han sido obtenidas mediante la implementación de los siguientes comandos:

```
> apc.plot.fit.pt(modelo_tasa_suicidios_males)
> apc.plot.fit.pt(modelo_tasa_suicidios_females)
```



(a) Representación de las transformadas de probabilidad para varones.



(b) Representación de las transformadas de probabilidad para mujeres.

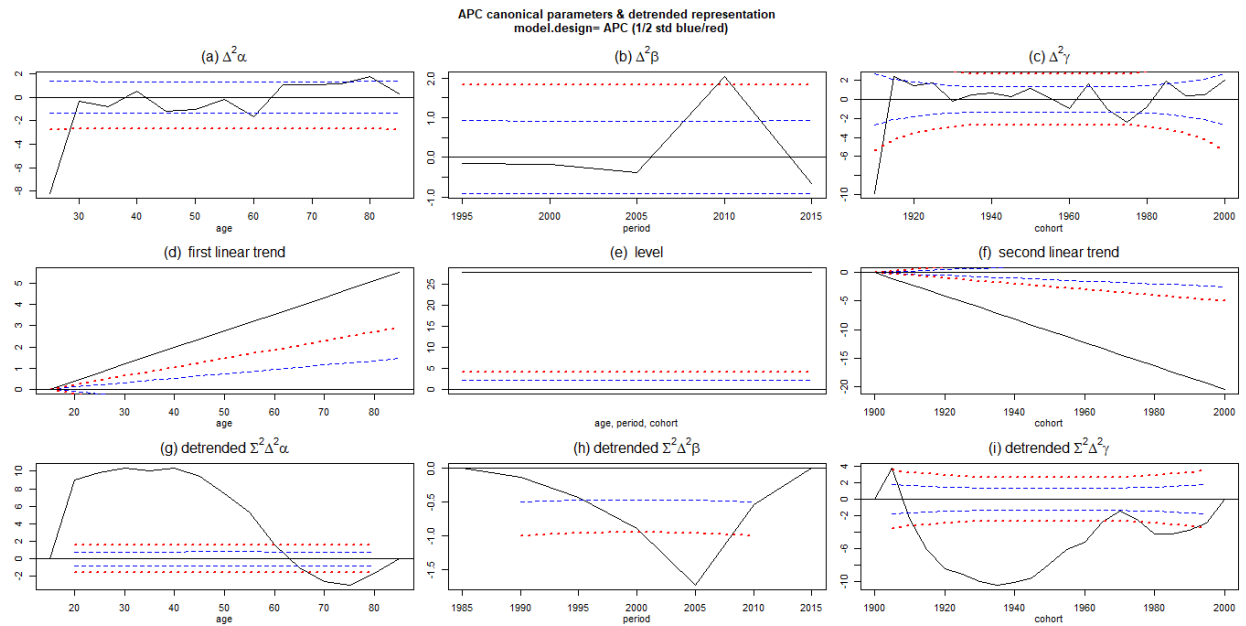
Figura 5.26: Representación de las transformadas de probabilidad diferenciando por sexos.

Al igual que en el caso general, representado en la figura 5.17, la mayoría de los datos se encuentran en el 80% de la distribución central. Además, con independencia del sexo, la mayoría de las observaciones que se encuentran en las colas pertenecen a los periodos 1985- 1989 y 1995- 1999; destacando, en este aspecto, los grupos con edades comprendidas entre los 15- 30 años y 60- 70 años.

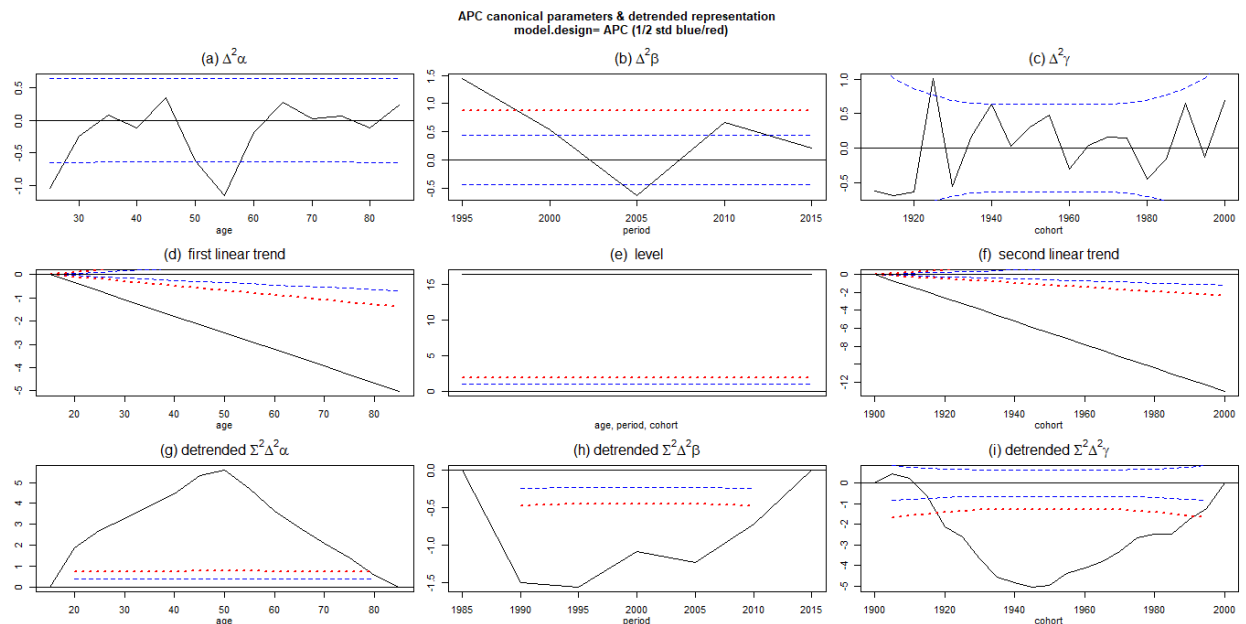
Por otro lado, las representaciones gráficas de las estimaciones de los modelos edad- periodo- cohorte, utilizando la parametrización canónica y el método *detrended*, se muestran en la figura 5.27. En este aspecto, dichas gráficas han sido obtenidas a partir de la implementación de los siguientes comandos:

```
> apc.plot.fit(modelo_tasa_suicidios_males,type="detrend")
> apc.plot.fit(modelo_tasa_suicidios_females,type="detrend")
```

En primer lugar, los paneles (a), (b) y (c) de las figuras 5.27a y 5.27b, muestran las estimaciones de las dobles diferencias; considerando, en este aspecto, cada uno de los regresores temporales del modelo. Por otro lado, las gráficas (d), (e) y (f) muestran el nivel y las pendientes asociadas al plano lineal. Finalmente, las



(a) Representación gráfica de las estimaciones de los parámetros haciendo uso de la metodología *detrended*. Segmento de varones.



(b) Representación gráfica de las estimaciones de los parámetros haciendo uso de la metodología *detrended*. Segmento de mujeres.

Figura 5.27: Representación gráfica de las estimaciones de los parámetros haciendo uso de la metodología *detrended*.

figuras (g), (h) e (i) muestran las partes no lineales asociadas a los efectos temporales del modelo; siendo obtenidas, estas últimas, mediante el cálculo de la dobles sumas de las dobles diferencias (evaluadas en base al método *detrended*). De tal forma, en relación con dichas gráficas, tanto el primer valor como el último, se encuentran fijados a cero. En este aspecto, en relación con las gráficas de las dobles sumas de las dobles diferencias empleando el método *detrended*; se observa que, existen diferencias entre ambos sexos. Dichas disimilitudes serán explicadas, de forma individual, a continuación:

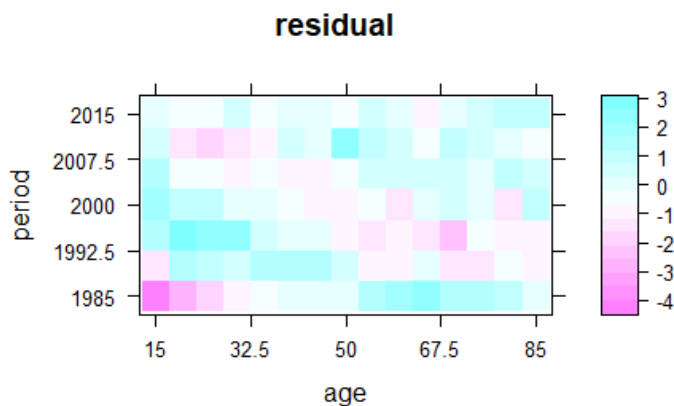
- **Gráfico (g).** Estas figuras representan las dobles sumas de las dobles diferencias referentes a la variable edad; empleando, a su vez, el método *detrended* ($\sum^2 \Delta^2 \alpha$). En primer lugar, en el caso de los varones, la relación entre la edad y las tasas de mortalidad por suicidio posee, entre los 15 y los 60 años, naturaleza cóncava; mientras que, entre los 60 y los 89 años, adquiere una forma convexa. Sin embargo, en el caso de las mujeres, la relación entre la edad y las tasas de mortalidad por suicidio posee forma de U invertida.
- **Gráfico (h).** Estas figuras representan las dobles sumas de las dobles diferencias concernientes a la variable periodo; empleando, a su vez, el método *detrended* ($\sum^2 \Delta^2 \beta$). Con independencia del sexo, a partir del periodo 2005- 2009, se produce un aumento de las tasas de mortalidad por suicidio; pudiendo estar motivado, este hecho, por el comienzo de la crisis económica de 2008 (*Gran Recesión*). Por otro lado, en el caso de las mujeres, la gráfica de las dobles sumas de las dobles diferencias muestra ciertas discontinuidades, respecto a la relación existente entre el regresor periodo y las tasas de mortalidad por suicidio; mientras que, en el caso de los varones, la tendencia de la relación existente entre dichos regresores posee un punto de inflexión en torno al periodo 2005- 2009.
- **Gráfico (i).** Estas figuras representan las dobles sumas de las dobles diferencias de la variable cohorte; empleando, a su vez, el método *detrended* ($\sum^2 \Delta^2 \gamma$). En primer lugar, en el caso del segmento de mujeres, la relación existente entre las tasas de mortalidad por suicidio y el regresor cohorte, posee forma de U. De tal forma, las mujeres nacidas a finales de la 2ª Guerra Mundial presentan menores tasas de mortalidad por suicidio, en comparación con las mujeres nacidas antes y después del conflicto. Por otro lado, en el caso de los varones, la relación existente entre las ratios de mortalidad por suicidio y el regresor cohorte posee distintos tramos. En este caso, los varones nacidos al comienzo de la 2ª Guerra Mundial poseen menores tasas de mortalidad por suicidio que aquellos nacidos antes y después del conflicto.

Finalmente, en la figura 5.28 se introducen las gráficas de los residuos de los modelos. Estas figuras han sido obtenidas mediante la implementación de los siguientes comandos:

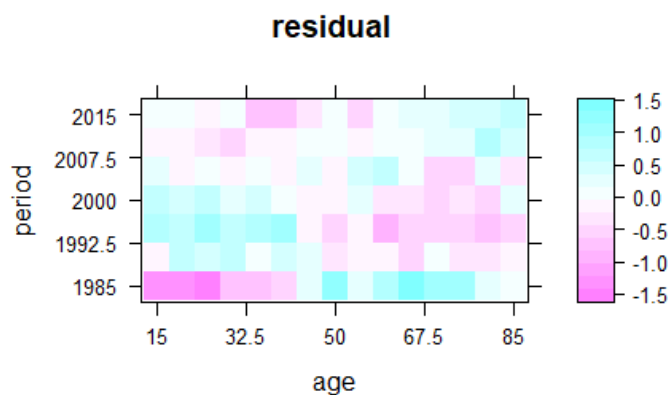
```
> apc.plot.fit.residuals(modelo_tasa_suicidios_males)
> apc.plot.fit.residuals(modelo_tasa_suicidios_females)
```

En primer lugar, considerando la leyenda situada en la parte derecha de ambas figuras; se observa que, los residuos asociados al modelo edad- periodo- cohorte aplicado a las tasas de mortalidad por suicidio de varones, poseen cifras superiores a los residuos obtenidos en el segmento de mujeres. Sin embargo, en ambos grupos, los residuos obtenidos al ajustar el modelo general son reducidos; concentrándose, de tal forma, en torno a cero. Por otro lado, los residuos que poseen mayores valores, se encuentran contenidos en el grupo de personas con edades comprendidas entre los 15 y los 29 años, independientemente del sexo considerado. Además, el periodo acaecido entre 1985 y 1989 presenta elevadas cifras de residuos, en comparación con el resto de periodos y con independencia del sexo. Esta última circunstancia puede estar motivada por las elevadas ratios de mortalidad por suicidio existentes en dicho periodo, tanto en el segmento de varones como en el de mujeres.

A pesar de estas circunstancias, el ajuste realizado a través de la implementación de un modelo edad- periodo- cohorte aplicado a las tasas de mortalidad por suicidio de varones y mujeres, presenta un buen comportamiento, al menos desde el punto de vista de los residuos obtenidos en ambos modelos.



(a) Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio de varones.



(b) Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio de mujeres.

Figura 5.28: Representación gráfica de los residuos obtenidos al implementar un modelo edad- periodo- cohorte a las tasas de mortalidad por suicidio.

5.4.4. Inferencia sobre el modelo

Finalmente, tras analizar los resultados obtenidos a través de la implementación de los modelos edad- periodo- cohorte aplicados, estos últimos, a las tasas de mortalidad por suicidio de varones y mujeres, se realizará la correspondiente inferencia sobre ambos modelos. Para ello, se utilizará la *tabla de deviance*; que se obtiene, en el caso del segmento de varones, a través de los siguientes comandos:

```
> deviance_table_men=apc.fit.table(tasa_suicidios_hombres,"gaussian.response")
> deviance_table_men
##      -2logL df.residual LR vs.APC df vs.APC prob(>chi_sq) F vs.APC prob(>F)
## APC 323.848         65      NaN      NaN      NaN      NaN      NaN
## AP  511.282         84  187.433      19      0.000  16.969      0.00
## AC  344.918         70   21.070       5      0.001   2.889      0.02
## PC  546.564         78  222.716      13      0.000  36.702      0.00
## Ad  521.607         89  197.759      24      0.000  15.101      0.00
## Pd  637.177         97  313.329      32      0.000  38.123      0.00
## Cd  551.020         83  227.172      18      0.000  27.812      0.00
## A   567.279         90  243.430      25      0.000  23.814      0.00
## P   640.732         98  316.884      33      0.000  38.309      0.00
```

```

## C 553.835      84 229.986      19      0.000 27.157 0.00
## t 640.398     102 316.550      37      0.000 34.053 0.00
## tA 657.402    103 333.553      38      0.000 39.287 0.00
## tP 643.848    103 320.000      38      0.000 34.322 0.00
## tC 649.662    103 325.814      38      0.000 36.373 0.00
## 1 660.343     104 336.495      39      0.000 39.414 0.00
##      aic
## APC 405.848
## AP 555.282
## AC 416.918
## PC 602.564
## Ad 555.607
## Pd 655.177
## Cd 597.020
## A 599.279
## P 656.732
## C 597.835
## t 648.398
## tA 663.402
## tP 649.848
## tC 655.662
## 1 664.343

```

De tal forma, si se consideran las columnas de $-2\log(L)$ (menos dos veces el logaritmo de la función de verosimilitud) y *aic* (criterio de información de Akaike) se obtiene que, el modelo edad- periodo- cohorte (columna APC), posee las menores cifras en relación con dichos estadísticos. Por tanto, en comparación con el resto de submodelos anidados, el modelo general es el que presenta un mejor ajuste. De tal forma, los regresores edad, periodo y cohorte resultan ser significativos a la hora de analizar las tasas de mortalidad por suicidio de varones.

Por otro lado, la *tabla de deviance*, en relación con el segmento de mujeres, se obtiene a través de los siguientes comandos:

```

> deviance_table_women=apc.fit.table(tasa_suicidios_mujeres,"gaussian.response")
> deviance_table_women
##      -2logL df.residual LR vs.APC df vs.APC prob(>chi_sq) F vs.APC prob(>F)
## APC 169.564      65      NaN      NaN      NaN      NaN      NaN
## AP 335.630      84 166.065      19      0 13.214      0
## AC 247.634      70 78.070      5      0 14.343      0
## PC 359.835      78 190.271      13      0 25.617      0
## Ad 366.915      89 197.351      24      0 15.032      0
## Pd 378.582      97 209.018      32      0 12.838      0
## Cd 382.892      83 213.328      18      0 23.931      0
## A 435.210      90 265.646      25      0 30.038      0
## P 421.142      98 251.578      33      0 19.656      0
## C 399.909      84 230.345      19      0 27.262      0
## t 400.367     102 230.803      37      0 14.068      0
## tA 453.984     103 284.420      38      0 23.966      0
## tP 436.157     103 266.593      38      0 19.956      0
## tC 421.572     103 252.008      38      0 17.146      0
## 1 476.887     104 307.323      39      0 29.449      0
##      aic
## APC 251.564
## AP 379.630
## AC 319.634
## PC 415.835

```

```

## Ad 400.915
## Pd 396.582
## Cd 428.892
## A 467.210
## P 437.142
## C 443.909
## t 408.367
## tA 459.984
## tP 442.157
## tC 427.572
## l 480.887

```

De tal forma, si se consideran las columnas de $-2\log(L)$ (menos dos veces el logaritmo de la función de verosimilitud) y aic (criterio de información de Akaike) se obtiene que, el modelo edad- periodo- cohorte (columna APC), posee las menores cifras en relación con dichos estadísticos. Por tanto, en comparación con el resto de submodelos anidados, el modelo general es el que presenta un mejor ajuste. De tal forma, los regresores edad, periodo y cohorte resultan ser significativos a la hora de analizar las tasas de mortalidad por suicidio de mujeres, al igual que ocurría en el segmento de los varones.

5.5. Discusión de los resultados

En primer lugar, el modelo general aplicado a las tasas de mortalidad por suicidio (por 100000 habitantes), generaba mejores estimaciones que el modelo edad- periodo- cohorte correspondiente a las cifras brutas de suicidios; como se observaba al comparar las gráficas de transformadas de probabilidad obtenidas mediante la implementación de dichos modelos. De tal forma, en el modelo general aplicado a las ratios de mortalidad por suicidio, la mayoría de los elementos correspondientes al gráfico de transformadas de probabilidad, se encontraban situados en el 80 % central de la distribución (figura 5.17); mientras que, en el modelo edad- periodo- cohorte aplicado a las cifras brutas de suicidios (figura 5.7), se obtuvo que, la mayoría de las estimaciones, se situaban en las colas de la distribución. Igualmente, los residuos arrojados por la estimación del modelo general aplicado a las tasas de mortalidad por suicidio (figura 5.19), presentaban una cuantía inferior a los residuos obtenidos mediante la implementación del modelo edad- periodo- cohorte aplicado a las cifras brutas de suicidios (figura 5.8); ilustrando el peor comportamiento, de este último modelo, a la hora de ajustar los datos.

Por otro lado, conviene tener en cuenta que, tanto en el modelo general aplicado a las tasas de mortalidad por suicidio como, en el modelo edad- periodo- cohorte implementado en base a las cifras brutas de suicidio; se obtuvo que, los tres regresores temporales del modelo (edad, periodo y cohorte), eran estadísticamente significativos.

Por otra parte, teniendo estas cuestiones en consideración, se realizará una discusión de los resultados en base al modelo que presentaba la mejor adecuación en relación con los datos; siendo, este último, el modelo edad- periodo- cohorte aplicado a las tasas de mortalidad por suicidio (considerando, de forma agregada, las cifras de hombres y mujeres). En este aspecto, realizar comparaciones entre los efectos de las variables edad, periodo y cohorte, para distintos años o regiones, reviste de una notable dificultad (posibles cambios en los sistemas de conteo de suicidios, diferencias regionales existentes en relación con la determinación de los suicidios, ...); por lo que, se debe ser muy cauto a la hora de realizar tales comparaciones (Granizo et al., 1996) [31].

En primer lugar, las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019) presentaban una tendencia creciente entre los 15 y los 49 años, edad a partir de la cual se producía una reducción en las ratios de mortalidad por suicidio. Finalmente, entre los 75 y los 89 años, se volvían a incrementar dichas tasas. De tal forma, en países como Corea del Sur y Japón (periodo 1986- 2015), también se producían incrementos, en relación con las tasas de mortalidad por suicidio, en el grupo de personas con edades comprendidas entre los 75 y los 89 años, como señalan autores como Kim et al. (2011) [12] y Kino et al. (2019) [3]. Además, autores como Granizo et al. (1996) [31], también encontraron un efecto positivo, del regresor

edad, sobre las tasas de mortalidad por suicidio. Sin embargo, estos resultados se contraponen, por ejemplo, con los obtenidos por Légaré y Hamel (2013) [7] en Quebec (periodo 1950- 2009); ya que, en esta provincia canadiense, las ratios de mortalidad por suicidio decrecían, de forma global, tras alcanzar, en edades adultas, el máximo de la serie.

Por otro lado, las tasas de mortalidad por suicidio de Inglaterra y Gales presentaron, entre 1985 y 2009, una tendencia decreciente; pudiendo estar motivado, este hecho, por la imposición de limitaciones, por parte del gobierno británico, en relación con el acceso a determinados métodos de suicidio. Estas conclusiones también fueron obtenidas por Kreitman (1976) [80] o Amos et al. (2001) [77]. Sin embargo, a partir de 2009, se produjo un incremento de las ratios de mortalidad por suicidio; pudiendo estar causado, este hecho, por el estallido de la crisis económica de 2008 y el posterior aumento del desempleo. De tal forma, Gunnell et al. (1999) [81] encontraron una relación importante entre los aumentos de las tasas de desempleo de Inglaterra y Gales y el incremento de las ratios de mortalidad por suicidio. Por otro lado, en otros países desarrollados como Corea del Sur o Japón, también se han encontrado evidencias estadísticas que relacionan los incrementos en las tasas de mortalidad por suicidio con los periodos de crisis económica (Kino et al., 2019) [3].

Por otra parte, las ratios de mortalidad por suicidio de las personas mayores de 50 años presentaron, entre 1985 y 2019, una tendencia decreciente, a diferencia de lo que ha ocurrido, por ejemplo, en Estados Unidos, donde las ratios de mortalidad por suicidio, pertenecientes a dicho grupo de edad, se han visto incrementadas durante las últimas décadas (Martínez- Alés et al., 2021) [4].

Por otra parte, las generaciones inglesas y galesas nacidas alrededor de 1940, presentaban menores tasas de mortalidad por suicidio que las generaciones nacidas entre finales de los años 40 y mediados de los años 70 del siglo XX (etapas posteriores al final de la 2ª Guerra Mundial). Estas conclusiones también fueron obtenidas por autores como Skegg y Cox (1991) [2], Granizo et al. (1996) [31], Beaupré y St- Laurent (1998) [41], Snowdon y Hunt (2002) [25], Légaré y Hamel (2013) [7], Phillips (2014) [42] y Kino et al. (2019) [3]; en relación, respectivamente, con las tasas de mortalidad por suicidio de Nueva Zelanda, España, Canadá, Australia, Quebec (provincia de Canadá), Estados Unidos y Japón.

Por otra parte, durante los últimos años de análisis, se observaba un aumento de las tasas de mortalidad por suicidio en los grupos de población más jóvenes. Estos últimos resultados fueron obtenidos, igualmente, por autores como Mojtabai et al. (2016) [55], Plemmons et al. (2018) [56], Keyes et al. (2019) [57], Raifman et al. (2020) [54] o Martínez- Alés et al. (2021) [4], investigadores que estudiaron las tasas de mortalidad por suicidio de Estados Unidos. De tal forma, según Oquendo y Volkow (2018) [58], la razón principal que subyace a este efecto de cohorte es la epidemia de opio que asola Estados Unidos; y que afecta, especialmente, a las cohortes más jóvenes. Por tanto, Inglaterra y Gales podrían estar sufriendo también este último problema, situación que podría explicar las elevadas tasas de mortalidad por suicidio existentes entre jóvenes, como señalan autores como Platt (2000) [84] o Gunnell et al. (2003) [78]. Además, las tasas de mortalidad por suicidio de jóvenes japoneses también se han visto incrementadas durante las últimas décadas; aunque, según autores como Abe (2012) [72] o Marchand et al. (2017) [71] dicho incremento está motivado, principalmente, por causas económicas.

Finalmente, el modelo edad- periodo- cohorte general presentaba el mejor ajuste, frente al resto de submodelos, en relación con las tasas de mortalidad por suicidio. De tal forma, los regresores edad, periodo y cohorte resultaban ser estadísticamente significativos a la hora de explicar dichas ratios. Estos resultados también fueron obtenidos por autores como Reed et al. (1985) [36], Barnes et al. (1986) [43], Mao et al. (1990) [44], Granizo et al. (1996) [31], Beaupré y St- Laurent (1998) [41], Légaré y Hamel (2013) [7] o Martínez- Alés et al. (2021) [4]. Sin embargo, la variable cohorte no resultaba ser estadísticamente significativa a la hora de explicar las tasas de mortalidad por suicidio de Inglaterra y Gales, correspondientes al periodo 1970-1980 (Murphy et al., 1986) [38]; a pesar de que, autores como Surtee y Duffy (1989) [86] y Charlton et al. (1993) [74], que también estudiaron las ratios de mortalidad por suicidio de Inglaterra y Gales; hallaron que, dicho regresor, resultaba ser significativo. Por otro lado, Kino et al. (2019) [3] encontraron que, la variable cohorte, resultaba ser no significativa a la hora de explicar las ratios de mortalidad por suicidio de Corea del Sur.

Por otro lado, se observaban diferencias importantes entre las tasas de mortalidad por suicidio de varones

y mujeres; coincidiendo, por tanto, con los resultados obtenidos por autores como Barnes et al. (1986) [43], Mao et al. (1990) [44], Granizo et al. (1996) [31], Beaupré y St- Laurent (1998) [41], Keyes y Li (2012) [6] o Légaré y Hamel (2013) [7], por señalar algunos ejemplos. Por tanto, se discutirán los resultados obtenidos a través de la aplicación de modelos edad- periodo- cohorte a los segmentos de varones y mujeres de Inglaterra y Gales (periodo 1985- 2019).

En primer lugar, las tasas de mortalidad por suicidio de varones fueron notablemente superiores, durante todo el periodo de análisis, a las ratios de mortalidad por suicidio de mujeres; coincidiendo, por tanto, con los resultados obtenidos por autores como Woodbury et al. (1988) [15] o Bille- Brahe (1993) [16]. En este aspecto, dicha característica se repite en la mayor parte de los estudios realizados en materia de suicidios.

Por otra parte, en relación con el regresor edad, las tasas de mortalidad por suicidio de varones presentaban una tendencia creciente entre los 15 y los 40 años, edad a partir de la cual se producía una disminución de las ratios de mortalidad por suicidio. Finalmente, entre los 70 y los 89 años, se volvía a producir un incremento en dichas tasas; alcanzando, en el grupo de edad que poseía entre 85 y 89 años, el máximo global de la serie, como también encontraron autores como Gunnell et al. (2003) [78]. Resultados similares fueron obtenidos, en el segmento de varones japoneses, por Kim et al. (2011) [12]. Por otro lado, considerando el segmento de mujeres, las tasas de mortalidad por suicidio se incrementaban entre los 15 y los 55 años (máximo global de la serie); disminuyendo, posteriormente entre los 56 y los 65 años. Finalmente, al igual que en el caso de los varones, las ratios de mortalidad por suicidio volvían a incrementarse en los grupos con mayor edad. Sin embargo, Granizo et al. (1996) [31] no encontraron diferencias significativas, en relación con la variable edad, entre las tasas de mortalidad por suicidio de hombres y mujeres.

Por otro lado, en relación con la variable periodo, las tasas de mortalidad por suicidio de varones se redujeron, de forma destacada, entre 1985 y 2009; incrementándose, posteriormente, a partir de dicha fecha. En este aspecto, autores como Jeon y Reither (2016) [68] o Kino et al. (2019) [3], encontraron evidencias estadísticas que relacionaban los aumentos en las ratios de mortalidad por suicidio de varones japoneses y coreanos, con los periodos de crisis económica. Sin embargo, en el caso del segmento de mujeres, las ratios de mortalidad por suicidio disminuyeron entre 1985 y 2015, año a partir del cual se produjo un ligero incremento en relación con dichas ratios.

Por otra parte, las cohortes que presentaban mayores tasas de mortalidad por suicidio, considerando el segmento de varones, se correspondían con aquellas nacidas entre 1945 y 1970 (etapas posteriores a la 2ª Guerra Mundial). Sin embargo, en el segmento de mujeres, las cohortes con mayores tasas de mortalidad por suicidio se correspondían con aquellas nacidas entre 1930 y 1970. Estos resultados también fueron obtenidos por autores como Murphy y Wetzel (1980) [34], Mao et al. (1990) [44], Skegg y Cox (1991) [2], Snowdon y Hunt (2002) [25] o Légaré y Hamel (2013) [7]; investigadores que estudiaron, respectivamente, las ratios de mortalidad por suicidio de Estados Unidos, Canadá, Nueva Zelanda, Australia y Quebec. Sin embargo, autores como Brinton (2010) [69] o Kino et al. (2019) [3] encontraron que, las tasas de mortalidad por suicidio de varones y mujeres japoneses nacidos en 1951, poseían menores tasas de mortalidad por suicidio que aquellos grupos de personas nacidos durante otros años.

Por otro lado, independientemente del sexo, las ratios de mortalidad por suicidio se incrementaban durante la vejez. Estas conclusiones también fueron obtenidas por autores como Khang et al. (2005) [61] o Kino et al. (2019) [3], que analizaron las tasas de mortalidad por suicidio de Corea del Sur y Japón. Sin embargo, en el caso de Quebec, las tasas de mortalidad por suicidio de hombres y mujeres se ven reducidas con la edad (Légaré y Hamel, 2013) [7].

Por otra parte, durante los últimos años de la serie, las ratios de mortalidad por suicidio de las mujeres con edades comprendidas entre los 15 y los 24 años, se han visto notablemente incrementadas. Estos resultados también fueron obtenidos por autores como Marchant et al. (2017) [71] o Martínez- Alés et al. (2021) [4], en relación con las tasas de mortalidad por suicidio, respectivamente, de Japón y Estados Unidos.

Finalmente, los regresores edad, periodo y cohorte eran significativos a la hora de explicar las tasas de mortalidad por suicidio de varones y mujeres, como también obtuvieron autores como Morrel et al. (2002)

[47], Odagiri et al. (2011) [48] o Légaré y Hamel (2013) [7]. Sin embargo, Kino et al. (2019) [3] encontraron que, la variable cohorte, resultaba no ser significativa a la hora de explicar las tasas de mortalidad por suicidio de varones y mujeres coreanos; mientras que, Skegg y Cox (1991) [2] hallaron que, la variable cohorte, resultaba ser no significativa a la hora de explicar las ratios de mortalidad por suicidio de mujeres neozelandesas.

Capítulo 6

Conclusiones y limitaciones del estudio

En este trabajo, se han estudiado los efectos que poseían las variables edad, periodo y cohorte a la hora de explicar las cifras brutas de suicidios y las tasas de mortalidad por suicidio (por 100000 habitantes), de Inglaterra y Gales, concernientes al periodo 1985- 2019. Los datos empleados han sido obtenidos de *Office for National Statistics*; poseyendo, en este aspecto, una estructura de *age- period array*, caracterizada por poseer intervalos de 5 años.

De tal forma, para estimar los modelos edad- periodo- cohorte correspondientes, se ha utilizado el paquete *apc*, desarrollado por Fannon y Nielsen (2020) [8]; y que se encuentra disponible, a su vez, en el software R. En este aspecto, dicho paquete utiliza la parametrización canónica propuesta por Kuang et al. (2008) [5], con el objetivo de solventar el problema de identificación asociado a este tipo de modelos.

Por otro lado, a la hora de estudiar las cifras brutas de suicidios se ha utilizado la familia de modelos Poisson; empleando, como grupo de exposición, a la población total. Sin embargo, a la hora de analizar las tasas de mortalidad por suicidio, se ha utilizado la familia de modelos gaussianas. En este aspecto, la estimación del modelo edad- periodo- cohorte concerniente a las ratios de mortalidad por suicidio, arrojaba mejores resultados que el modelo general aplicado a las cifras brutas de suicidio, como quedaba patente al comparar las gráficas de transformadas de probabilidad y los residuos arrojados por ambos modelos. Partiendo del mejor comportamiento del modelo general aplicado a las tasas de mortalidad por suicidio, se realizó una segmentación en función del sexo; motivada, esta última, por las enormes diferencias que existían, en materia de suicidio, entre varones y mujeres.

De tal forma, en cuanto a las conclusiones obtenidas, las tasas de mortalidad por suicidio de Inglaterra y Gales presentaban una tendencia creciente entre los 15 y los 49 años, edad a partir de la cual se producía una reducción de dichas ratios. Posteriormente, entre los 75 y los 89 años, las tasas de mortalidad por suicidio volvían a incrementarse, coincidiendo con los resultados obtenidos, para Corea del Sur y Japón, por Kim et al. (2011) [12] y Kino et al. (2019) [3]. En este aspecto, en el caso de los varones, el grupo de edad que poseía entre 85 y 89 años presentaba el máximo global de la serie, como también encontraron autores como Gunnell et al. (2003) [78]. Sin embargo, las mujeres que poseían entre 51 y 55 años, presentaron las mayores tasas de mortalidad por suicidio, considerando el periodo de análisis (1985- 2019).

Por otro lado, las tasas de mortalidad por suicidio de varones poseían una tendencia decreciente entre 1985 y 2009; coincidiendo, este último año, con un periodo de crisis económica. De tal forma, estos hechos parecen confirmar las ideas planteadas por autores como Gunnell et al. (1999) [81], que encontraron una importante relación directa entre el incremento del desempleo y el aumento de las tasas de mortalidad por suicidio de Inglaterra y Gales. Además, la reducción acaecida, en materia de suicidios, entre 1985 y 2009, podría estar motivada por los mayores controles, implementados por el gobierno británico, en relación con el acceso a ciertos instrumentos de suicidio. Sin embargo, en el segmento de mujeres, las tasas de mortalidad por suicidio se redujeron entre 1985 y 2015; incrementándose, ligeramente, a partir de dicha fecha.

Por otra parte, en relación con el segmento de mujeres, las cohortes con mayores tasas de mortalidad por suicidio se correspondían con aquellas nacidas entre 1930 y 1970; mientras que, en el caso de los varones, las mayores ratios de mortalidad por suicidio pertenecían a las cohortes de hombres nacidas entre 1945 y 1970. Por tanto, estas conclusiones coinciden con las obtenidas por autores como Murphy y Wetzel (1980) [34], Mao et al. (1990) [44], Skegg y Cox (1991) [2], Snowdon y Hunt (2002) [25] o Légaré y Hamel (2013) [7].

En relación con la vejez, las tasas de mortalidad por suicidio tendían a incrementarse, independientemente del sexo, en los grupos de personas con edades más avanzadas, coincidiendo con los resultados obtenidos por Khang et al. (2005) [61] y Kino et al. (2019) [3]. Además, durante los últimos años de la serie, las tasas de mortalidad por suicidio de los grupos de población más jóvenes se han visto incrementadas; considerando, principalmente, el segmento de mujeres. Estos resultados también fueron obtenidos por Marchant et al. (2017) [71] o Martínez- Alés et al. (2021) [4]; en relación, respectivamente, con las ratios de mortalidad por suicidio de Japón y Estados Unidos.

Por otro lado, los regresores edad, periodo y cohorte resultaban ser estadísticamente significativos en todas las estimaciones realizadas; ya que, los modelos edad- periodo- cohorte presentaban el mejor ajuste posible, frente al resto de submodelos anidados. Por tanto, estos últimos resultados coinciden con los obtenidos por autores como Reed et al. (1985) [36], Barnes et al. (1986) [43], Mao et al. (1990) [44], Granizo et al. (1996) [31], Beaupré y St- Laurent (1998) [41], Morrel et al. (2002) [47], Odagiri et al. (2011) [48], Légaré y Hamel (2013) [7] y Martínez- Alés et al. (2021) [4].

Finalmente, en cuanto a las limitaciones de este estudio conviene tener en que, las estadísticas utilizadas en el actual proyecto, están basadas en el año de registro del suicidio (no en el año en el que este se produjo); dificultado, esta circunstancia, las tareas de análisis de tendencia. Además, las comparaciones realizadas en el apartado de discusión de resultados, deben ser tomadas con cautela, debido a las diferencias que existen, desde un punto de vista temporal y geográfico, en relación con los sistemas de conteo de los suicidios.

Por otro lado, como extensión de este análisis, podrían considerarse distintos subgrupos, divididos en función de la edad de los individuos; y estimar, en base a dichos grupos, los modelos edad- periodo- cohorte correspondientes. Por otro lado, los problemas psicológicos y económicos derivados de la pandemia de COVID- 19, pueden haber incrementado el número de suicidios; por lo que, el análisis de dichas estadísticas, podría generar información relevante a la hora de tomar decisiones relacionadas con las políticas de salud pública. Finalmente, partiendo de las elevadas divergencias que existen, en materia de suicidios, entre las diversas regiones que componen Inglaterra y Gales (ver figuras 2.4 y 2.5), convendría realizar un análisis en el que se considerasen dichas diferencias geográficas.

Bibliografía

- [1] Bell, A., & Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social science & medicine*, 93, 163-165.
- [2] Skegg, K., & Cox, B. (1991). Suicide in New Zealand 1957–1986: the influence of age, period and birth-cohort. *Australian and New Zealand Journal of Psychiatry*, 25(2), 181-190.
- [3] Kino, S., Jang, S. N., Gero, K., Kato, S., & Kawachi, I. (2019). Age, period, cohort trends of suicide in Japan and Korea (1986–2015): A tale of two countries. *Social Science & Medicine*, 235, 112385.
- [4] Martínez-Alés, G., Pamplin, J. R., Rutherford, C., Gimbrone, C., Kandula, S., Olfson, M., ... & Keyes, K. M. (2021). Age, period, and cohort effects on suicide death in the United States from 1999 to 2018: moderation by sex, race, and firearm involvement. *Molecular psychiatry*, 1-9.
- [5] Kuang, D., Nielsen, B., & Nielsen, J. P. (2008). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika*, 95(4), 979-986.
- [6] Keyes, K.M., Li G. (2012). Age-period-cohort modeling. In: Li G, Baker SP (Eds.), *Injury Research Theory, Methods, and Approaches*. New York: Springer, 409-26.
- [7] Légaré, G., & Hamel, D. (2013). An age-period-cohort approach to analyzing trends in suicide in Quebec between 1950 and 2009. *Canadian journal of public health*, 104(2), e118-e123.
- [8] Fannon, Z. & Nielsen, B. (2020). Package apc. Version 2.0.0. Available online: <https://cran.r-project.org/web/packages/apc/apc.pdf>.
- [9] Naghavi, M. (2019). Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *bmj*, 364. Available online: <https://www.bmj.com/content/364/bmj.194>
- [10] Diekstra, R. F. (1993). The epidemiology of suicide and parasuicide. *Acta Psychiatrica Scandinavica*, 87(S371), 9-20.
- [11] La Vecchia, C., Lucchini, F., & Levi, F. (1994). Worldwide trends in suicide mortality, 1955–1989. *Acta Psychiatrica Scandinavica*, 90(1), 53-64.
- [12] Kim, S. Y., Kim, M. H., Kawachi, I., & Cho, Y. (2011). Comparative epidemiology of suicide in South Korea and Japan: effects of age, gender and suicide methods. *Crisis*. Available online: <https://econtent.hogrefe.com/doi/abs/10.1027/0227-5910/a000046?journalCode=cric>
- [13] Sainsbury, P., & Jenkins, J. S. (1982). The accuracy of officially reported suicide statistics for purposes of epidemiological research. *Journal of Epidemiology & Community Health*, 36(1), 43-48.
- [14] Moens, G. F. (1985). The reliability of reported suicide mortality statistics: an experience from Belgium. *International journal of epidemiology*, 14(2), 272-275.
- [15] Woodbury, M. A., Manton, K. G., & Blazer, D. (1988). Trends in US suicide mortality rates 1968 to 1982: Race and sex differences in age, period and cohort components. *International Journal of Epidemiology*, 17(2), 356-362.

- [16] Bille-Brahe, U. (1993). The role of sex and age in suicidal behavior. *Acta Psychiatrica Scandinavica*, 87(S371), 21-27.
- [17] Moens, G. F., Haenen, W., & Van De Voorde, H. (1988). Epidemiological aspects of suicide among the young in selected European countries. *Journal of Epidemiology & Community Health*, 42(3), 279-285.
- [18] Coté, T. R., Biggar, R. J., & Dannenberg, A. L. (1992). Risk of suicide among persons with AIDS: a national assessment. *Jama*, 268(15), 2066-2068. Available online: <https://jamanetwork.com/journals/jama/article-abstract/400638>
- [19] Värnik, P., Sisask, M., Värnik, A., Laido, Z., Meise, U., Ibelshäuser, A., ... & Hegerl, U. (2010). Suicide registration in eight European countries: A qualitative analysis of procedures and practices. *Forensic Science International*, 202(1-3), 86-92.
- [20] Rockett, I. R., Kapusta, N. D., & Bhandari, R. (2011). Suicide misclassification in an international context: revisitation and update. *Suicidology online*, 2, 48-61.
- [21] Tøllefsen, I. M., Hem, E., & Ekeberg, Ø. (2012). The reliability of suicide statistics: a systematic review. *BMC psychiatry*, 12(1), 1-11.
- [22] Aouba, A., Joubla, E., Camelin, L., & Pequitnot, F. (2011). Évaluation de la qualité et amélioration de la connaissance des données de mortalité par suicide en France métropolitaine, 2006. *Bulletin épidémiologique hebdomadaire*, (47-48), 497-500.
- [23] Charron, M. F. (1982). La prévention du suicide au Québec: les indications dégagées par une étude statistique. *Santé mentale au Québec*, 7(2), 160-175.
- [24] St-Laurent, D., & Bouchard, C. (2004). L'épidémiologie du suicide au Québec: que savons-nous de la situation récente?. *Québec: Institut national de santé publique du Québec*.
- [25] Snowdon, J., & Hunt, G. E. (2002). Age, period and cohort effects on suicide rates in Australia, 1919–1999. *Acta Psychiatrica Scandinavica*, 105(4), 265-270.
- [26] Biddle, L., Brock, A., Brookes, S. T., & Gunnell, D. (2008). Suicide rates in young men in England and Wales in the 21st century: time trend study. *bmj*, 336(7643), 539-542. Available online: <https://www.bmj.com/content/336/7643/539.short>
- [27] Hawton, K., & Van Heeringen, K. (Eds.). (2000). The international handbook of suicide and attempted suicide. *John Wiley & Sons*.
- [28] Gagné, M., Robitaille, Y., Hamel, D., & St-Laurent, D. (2010). Firearms regulation and declining rates of male suicide in Quebec. *Injury Prevention*, 16(4), 247-253.
- [29] Hanigan, I. C., Butler, C. D., Kokic, P. N., & Hutchinson, M. F. (2012). Suicide and drought in new South Wales, Australia, 1970–2007. *Proceedings of the National Academy of Sciences*, 109(35), 13950-13955.
- [30] Chang, S. S., Gunnell, D., Sterne, J. A., Lu, T. H., & Cheng, A. T. (2009). Was the economic crisis 1997–1998 responsible for rising suicide rates in East/Southeast Asia? A time–trend analysis for Japan, Hong Kong, South Korea, Taiwan, Singapore and Thailand. *Social science & medicine*, 68(7), 1322-1331.
- [31] Granizo, J. J., Guallar, E., & Rodriguez-Artalejo, F. (1996). Age-period-cohort analysis of suicide mortality rates in Spain, 1959–1991. *International journal of epidemiology*, 25(4), 814-820.
- [32] Kupper, L. L., Janis, J. M., Karmous, A., & Greenberg, B. G. (1985). Statistical age-period-cohort analysis: a review and critique. *Journal of chronic diseases*, 38(10), 811-830.
- [33] Frome, E. L., & Checkoway, H. (1985). Use of Poisson regression models in estimating incidence rates and ratios. *American journal of epidemiology*, 121(2), 309-323.

- [34] Murphy, G. E., & Wetzel, R. D. (1980). Suicide risk by birth cohort in the United States, 1949 to 1974. *Archives of General Psychiatry*, 37(5), 519-523.
- [35] Häfner, H., & Schmidtke, A. (1985). Do cohort effects influence suicide rates?. *Archives of general psychiatry*, 42(9), 926-927.
- [36] Reed, J., Camus, J., & Last, J. M. (1985). Suicide in Canada: Birth-cohort analysis Le Suicide au Canada: analyse par cohorte de naissance. *Canadian Journal of Public Health*, 76(1), 43-47.
- [37] La Vecchia, C., Bollini, P., Imazio, C., & Decarli, A. (1986). Age, period of death and birth cohort effects on suicide mortality in Italy, 1955–1979. *Acta Psychiatrica Scandinavica*, 74(2), 137-143.
- [38] Murphy, E., Lindsay, J., & Grundy, E. (1986). 60 years of suicide in England and Wales: a cohort study. *Archives of General Psychiatry*, 43(10), 969-976.
- [39] Moens, G. F. G., Van Oortmarsen, G. J., Honggokoesoemo, S., & Van de Voorde, H. (1987). Birth cohort analysis of suicide mortality in Belgium 1954–1981 by a graphic and a quantitative method. *Acta Psychiatrica Scandinavica*, 76(4), 450-455.
- [40] Åsgård, U., Nordström, P., & Råbäck, G. (1987). Birth cohort analysis of changing suicide risk by sex and age in Sweden 1952 to 1981. *Acta Psychiatrica Scandinavica*, 76(4), 456-463.
- [41] Beaupré, M., & St-Laurent, D. (1998). Deux causes de décès: le cancer et le suicide. D'une génération à l'autre: évolution des conditions de vie (pp. 47-81). *Sainte-Foy: Bureau de la Statistique du Québec*.
- [42] Phillips, J. A. (2014). A changing epidemiology of suicide? The influence of birth cohorts on suicide rates in the United States. *Social Science & Medicine*, 114, 151-160.
- [43] Barnes, R. A., Ennis, J., & Schober, R. (1986). Cohort analysis of Ontario suicide rates, 1877–1976. *The Canadian Journal of Psychiatry*, 31(3), 208-213.
- [44] Mao, Y., Hasselback, P., Davies, J. W., Nichol, R., & Wigle, D. T. (1990). Suicide in Canada: An Epidemiological Assessment. *Canadian journal of public health*, 81(4), 324-328.
- [45] Keyes, K. M., Utz, R. L., Robinson, W., & Li, G. (2010). What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. *Social science & medicine*, 70(7), 1100-1108.
- [46] Boyer, R., & St-Laurent, D. (1999). La mortalité par suicide au Québec. *Frontières*, 12(1), 50-53.
- [47] Morrell, S., Page, A., & Taylor, R. (2002). Birth cohort effects in New South Wales suicide, 1865–1998. *Acta Psychiatrica Scandinavica*, 106(5), 365-372.
- [48] Odagiri, Y., Uchida, H., & Nakano, M. (2011). Gender differences in age, period, and birth-cohort effects on the suicide mortality rate in Japan, 1985-2006. *Asia Pacific Journal of Public Health*, 23(4), 581-587.
- [49] Stone, D. M., Simon, T. R., Fowler, K. A., Kegler, S. R., Yuan, K., Holland, K. M., ... & Crosby, A. E. (2018). Vital signs: trends in state suicide rates—United States, 1999–2016 and circumstances contributing to suicide—27 states, 2015. *Morbidity and Mortality Weekly Report*, 67(22), 617.
- [50] Wang, Z., Yu, C., Wang, J., Bao, J., Gao, X., & Xiang, H. (2016). Age-period-cohort analysis of suicide mortality by gender among white and black Americans, 1983–2012. *International journal for equity in health*, 15(1), 1-9.
- [51] Smith, T. W., & Son, J. (2019). General Social Survey final report: Trends in gun ownership in the United States, 1972–2018. Available online: <https://gssdataexplorer.norc.org/documents/905/download>
- [52] Ludwig, J., & Cook, P. J. (2000). Homicide and suicide rates associated with implementation of the Brady Handgun Violence Prevention Act. *Jama*, 284(5), 585-591. Available online: <https://jamanetwork.com/journals/jama/article-abstract/192946>

- [53] Andrés, A. R., & Hempstead, K. (2011). Gun control and suicide: The impact of state firearm regulations in the United States, 1995–2004. *Health Policy*, 101(1), 95-103.
- [54] Raifman, J., Larson, E., Barry, C. L., Siegel, M., Ulrich, M., Knopov, A., & Galea, S. (2020). State handgun purchase age minimums in the US and adolescent suicide rates: regression discontinuity and difference-in-differences analyses. *bmj*, 370. Available online: <https://www.bmj.com/content/370/bmj.m2436>
- [55] Mojtabai, R., Olfson, M., & Han, B. (2016). National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*, 138(6).
- [56] Plemmons, G., Hall, M., Doupnik, S., Gay, J., Brown, C., Browning, W., ... & Williams, D. (2018). Hospitalization for suicide ideation or attempt: 2008–2015. *Pediatrics*, 141(6).
- [57] Keyes, K. M., Gary, D., O'Malley, P. M., Hamilton, A., & Schulenberg, J. (2019). Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. *Social psychiatry and psychiatric epidemiology*, 54(8), 987-996.
- [58] Oquendo, M. A., & Volkow, N. D. (2018). Suicide: a silent contributor to opioid-overdose deaths. *New England journal of medicine*, 378(17), 1567-1569.
- [59] Motohashi, Y. (1991). Effects of socioeconomic factors on secular trends in suicide in Japan, 1953–86. *Journal of Biosocial Science*, 23(2), 221-227.
- [60] Suzuki, E., Kashima, S., Kawachi, I., & Subramanian, S. V. (2014). Prefecture-level economic conditions and risk of suicide in Japan: A repeated cross-sectional analysis 1975–2010. *The European Journal of Public Health*, 24(6), 949-954.
- [61] Khang, Y. H., Lynch, J. W., & Kaplan, G. A. (2005). Impact of economic crisis on cause-specific mortality in South Korea. *International journal of epidemiology*, 34(6), 1291-1301.
- [62] Jung-Choi, K., & Kim, Y. M. (2013). Socioeconomic inequalities in health status in Korea. *Journal of the Korean Medical Association*, 56(3), 167-174.
- [63] Park, S., Song, Y. J., Kim, J., Ki, M., Shin, J. Y., Kwon, Y. M., & Lim, J. (2018). Age, period, and cohort effects on suicide mortality in South Korea, 1992–2015. *International journal of environmental research and public health*, 15(8), 1580.
- [64] Lee, W. Y., Khang, Y. H., Noh, M., Ryu, J. I., Son, M., & Hong, Y. P. (2009). Trends in educational differentials in suicide mortality between 1993-2006 in Korea. *Yonsei medical journal*, 50(4), 482-492.
- [65] Yoon, T. H., Noh, M., Han, J., Jung-Choi, K., & Khang, Y. H. (2015). Deprivation and suicide mortality across 424 neighborhoods in Seoul, South Korea: a Bayesian spatial analysis. *International journal of public health*, 60(8), 969-976.
- [66] Yang, Y., & Land, K. C. (2013). Age-period-cohort analysis: New models, methods, and empirical applications. *Taylor & Francis*. Available online: <https://library.oapen.org/handle/20.500.12657/41269>
- [67] Kondo, N., & Oh, J. (2010). Suicide and karoshi (death from overwork) during the recent economic crises in Japan: the impacts, mechanisms and political responses. *Journal of Epidemiology & Community Health* 2010;64:649-650.
- [68] Jeon, S. Y., Reither, E. N., & Masters, R. K. (2016). A population-based analysis of increasing rates of suicide mortality in Japan and South Korea, 1985–2010. *BMC public health*, 16(1), 1-9.
- [69] Brinton, M. C. (2010). *Lost in transition: Youth, work, and instability in postindustrial Japan*. Cambridge University Press.
- [70] Vaughn, M. S., Huang, F. F., & Ramirez, C. R. (1995). DRUG abuse and ANTI-DRUG policy IN Japan: past history and future directions. *The British Journal of Criminology*, 35(4), 491-524.

- [71] Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., ... & John, A. (2017). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. *PLoS one*, 12(8).
- [72] Abe, A. (2012). Poverty and social exclusion of women in Japan. *Japanese Journal of Social Security Policy*, 9(1), 61-82.
- [73] Goldney, R. D., & Katsikitis, M. (1983). Cohort analysis of suicide rates in Australia. *Archives of General Psychiatry*, 40(1), 71-74.
- [74] Charlton, J., Kelly, S., Dunnell, K., Evans, B., Jenkins, R., & Wallis, R. (1993). Trends in suicide deaths in England and Wales. *Population Trends*, no.69, (1993), p.10-16
- [75] McClure, G. M. G. (2000). Changes in suicide in England and Wales, 1960–1997. *The British Journal of Psychiatry*, 176(1), 64-67.
- [76] Surtees, P. G., & Duffy, J. C. (1989). Suicide in England and Wales 1946–1985: an age-period-cohort analysis. *Acta Psychiatrica Scandinavica*, 79(3), 216-223.
- [77] Amos, T., Appleby, L., & Kiernan, K. (2001). Changes in rates of suicide by car exhaust asphyxiation in England and Wales. *Psychological medicine*, 31(5), 935-939.
- [78] Gunnell, D., Middleton, N., Whitley, E., Dorling, D., & Frankel, S. (2003). Influence of cohort effects on patterns of suicide in England and Wales, 1950–1999. *The British Journal of Psychiatry*, 182(2), 164-170.
- [79] Allebeck, P., Brandt, L., Nordstrom, P., & Åsgård, U. (1996). Are suicide trends among the young reversing? Age, period and cohort analyses of suicide rates in Sweden. *Acta Psychiatrica Scandinavica*, 93(1), 43-48.
- [80] Kreitman, N. (1976). The coal gas story. United Kingdom suicide rates, 1960-71. *Journal of Epidemiology & Community Health*, 30(2), 86-93.
- [81] Gunnell, D., Lopatzidis, T., Dorling, D., Wehner, H., Southall, H., & Frankel, S. (1999). Suicide and unemployment in young people: analysis of trends in England and Wales, 1921–1995. *The British Journal of Psychiatry*, 175(3), 263-270.
- [82] Christophersen, O., Rooney, C., & Kelly, S. (1998). Drug-related mortality: methods and trends. *Population trends-London-*, 29-37.
- [83] Agid, O., Shapira, B., Zislin, J., Ritsner, M., Hanin, B., Murad, H., ... & Lerer, B. (1999). Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar disorder and schizophrenia. *Molecular psychiatry*, 4(2), 163-172.
- [84] Platt, S. (2000). The Labour Market. The international handbook of suicide and attempted suicide, 309. *John Wiley & Sons Ltd.*
- [85] Pounder, D. J. (1993). Why are the British hanging themselves?. *The American journal of forensic medicine and pathology*, 14(2), 135-140.
- [86] Surtees, P. G., & Duffy, J. C. (1989). Suicide in England and Wales 1946–1985: an age-period-cohort analysis. *Acta Psychiatrica Scandinavica*, 79(3), 216-223.
- [87] Martinez-Miranda, M. D., Nielsen, B., & Nielsen, J. P. (2014). Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 178(1), 29-55.
- [88] Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in medicine*, 25(6), 977-993.
- [89] Mason, K.O., W.M. Mason, H.H. Winsborough, and W.K. Poole. 1973. Some method-ological issues in cohort analysis of archival data. *American Sociological Review*, 38:242–58.

- [90] Osmond, C., & Gardner, M. J. (1989). Age, period, and cohort models. Non-overlapping cohorts don't resolve the identification problem. *American Journal of Epidemiology*, 129(1), 31-35.
- [91] Fannon, Z., & Nielsen, B. (2018). Age-period cohort models (No. 2018-W04). In *Oxford Research Encyclopedia of Economics and Finance*.
- [92] Bell, F. C., & Miller, M. L. (2005). Life tables for the United States social security area, 1900-2100. *Social Security Administration, Office of the Chief Actuary* (No. 120).
- [93] Stockard, J., and R.M. O'Brien. 2002. Cohort effects on suicide rates: International variations. *American Sociological Review*, 67:854-72.
- [94] Comstock, G. W. (2001). Cohort analysis: WH Frost's contributions to the epidemiology of tuberculosis and chronic disease. *Sozial-und Präventivmedizin*, 46(1), 7-12.
- [95] McKenzie, D. J. (2006). Disentangling age, cohort and time effects in the additive model. *Oxford bulletin of economics and statistics*, 68(4), 473-495.
- [96] Attanasio, O. P. (1998). Cohort analysis of saving behavior by US households. *Journal of Human Resources*, 575-609.
- [97] Browning, M., Crossley, T. F., & Lührmann, M. (2016). Durable purchases over the later life cycle. *Oxford Bulletin of Economics and Statistics*, 78(2), 145-169.
- [98] Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in medicine*, 26(15), 3018-3045.
- [99] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591-605.
- [100] Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, 311-324.
- [101] Nielsen, B., & Nielsen, J. P. (2014). Identification and forecasting in mortality models. *The Scientific World Journal*, 2014.
- [102] Ejrnæs, M., & Hochguertel, S. (2013). Is business failure due to lack of effort? Empirical evidence from a large administrative sample. *The Economic Journal*, 123(571), 791-830.
- [103] Schulhofer-Wohl, S. (2018). The age-time-cohort problem and the identification of structural parameters in life-cycle models. *Quantitative Economics*, 9(2), 643-658.
- [104] O'Brien, R. (2014). Age-period-cohort models: approaches and analyses with aggregate data. *CRC Press*.
- [105] O'Brien, R. M. (2011). Constrained estimators and age-period-cohort models. *Sociological Methods & Research*, 40(3), 419-452.
- [106] Mazumdar, S., Li, C. C., & Bryce, G. R. (1980). Correspondence between a linear restriction and a generalized inverse in linear model analysis. *The American Statistician*, 34(2), 103-105.
- [107] O'Brien, R. M. (2003). UCR violent crime rates, 1958-2000: recorded and offender-generated trends. *Social Science Research*, 32(3), 499-518.
- [108] Meghir, C., & Whitehouse, E. (1996). The evolution of wages in the United Kingdom: Evidence from micro data. *Journal of labor Economics*, 14(1), 1-25.
- [109] Voas, D., & Chaves, M. (2016). Is the United States a counterexample to the secularization thesis?. *American Journal of Sociology*, 121(5), 1517-1556.
- [110] Poirier, D. J. (1998). Revising beliefs in nonidentified models. *Econometric Theory*, 483-509.

- [111] Berzuini, C., & Clayton, D. (1994). Bayesian analysis of survival on multiple time scales. *Statistics in medicine*, 13(8), 823-838.
- [112] Smith, T. R., & Wakefield, J. (2016). A review and comparison of age–period–cohort models for cancer incidence. *Statistical Science*, 31(4), 591-610.
- [113] Strang, G. 1998. Introduction to Linear Algebra (2nd edition). *Wellesley, MA: Wellesley- Cambridge Press*.
- [114] Rodgers, W.L. 1982. Estimable functions of age, period, and cohort effects. *American Sociological Review*, 47:774–87.
- [115] Yang, Y., S. Schulhofer- Wohl, W.J. Fu, and K.C. Land. 2008. The intrinsic estimator for age- period-cohort analysis: What it is and how to use it. *American Journal of Sociology*, 113:1697–73.
- [116] Searle, S.R. 1971. Linear Models. *New York: John Wiley & Sons*.
- [117] O’Brien, R. M. (2012). Visualizing rank deficient models: a row equation geometry of rank deficient matrices and constrained-regression. *PloS one*, 7(6), e38923. Available online: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0038923>
- [118] Yang, Y., S. Schulhofer- Wohl, W.J. Fu, and K.C. Land. 2008. The intrinsic estimator for age- period-cohort analysis: What it is and how to use it. *American Journal of Sociology*, 113:1697–736.
- [119] Holford, T.R. 1985. An alternative approach to statistical age- period- cohort analysis. *Journal of Chronic Disease*, 38:831 –6.
- [120] Yang, Y., and K.C. Land. 2006. A mixed models approach to the age- period- cohort analysis of repeated cross- section surveys, with applications to data on trends in verbal test scores. *Sociological Methodology*, 36:75–9.
- [121] Snijders, T.A.B. 2005. Fixed and random effects. In *Encyclopedia of Statistics in Behavioral Science*, vol. 2, eds. *B.S. Everitt and D.C. Howell*, 664–5. *Chichester: Wiley*.
- [122] Luo, L., and J. Hodges. 2013. The cross- classified age- period- cohort model as a constrained estimator. Presented at the annual meeting of the Population Association of America, April. Available online: <https://paa2013.princeton.edu/papers/132093>
- [123] O’Brien, R.M., J. Stockard, and L. Isaacson. 1999. The enduring effects of cohort size and percent of nonmarital births on age- specific homicide rates, 1960–1995. *American Journal of Sociology*, 104:1061–95.
- [124] Winship, C., and D.J. Harding. 2008. A mechanism based approach to the identification of age- period- cohort models. *Sociological Methods & Research*, 36:362–401.
- [125] Carstensen, B. & Plummer, M. & Laara, E. & Hills M. (2021). Package Epi. Version 2.44. Available in: <https://cran.r-project.org/web/packages/Epi/Epi.pdf>.
- [126] Schulhofer-Wohl, S., & Yang, Y. (2006). APC: Stata module for estimating age-period-cohort effects. Available online: <https://econpapers.repec.org/software/bocbocode/s456754.htm>
- [127] Sasieni, P. D. (2012). Age–period–cohort models in Stata. *The Stata Journal*, 12(1), 45-60.
- [128] Chauvel, L. (2012). APCD: Stata module for estimating age-period-cohort effects with detrended coefficients. Available online: <https://econpapers.repec.org/software/bocbocode/s457440.htm>
- [129] Nielsen, B. (2015). *apc: An R Package for Age-Period-Cohort Analysis*. *R J.*, 7(2), 52. Available online: https://www.nuffield.ox.ac.uk/economics/Papers/2014/nielsen_apc_package.pdf
- [130] De Angelis, D. and Gilks, W.R. (1994) Estimating acquired immune deficiency syndrome incidence accounting for reporting delay. *Journal of the Royal Statistical Society A*, 157, 31-40.

- [131] Clayton, D. and Schifflers, E. (1987) Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Statistics in Medicine*, 6, 449-467.
- [132] Jacobsen, R, von Euler, M, Osler, M, Lynge, E and Keiding, N (2004). Women's death in Scandinavia - what makes Denmark different? *European Journal of Epidemiology* 19, 117-121.
- [133] Riebler, A and Held, L. (2010) The analysis of heterogeneous time trends in multivariate age- period-cohort models. *Biostatistics*, 11, 57-59.

Apéndice A

Código en R

En este apéndice se introducirá el código que ha sido implementado, en R, a la hora de analizar las cifras y las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019). La descripción de los resultados obtenidos a partir de la ejecución de estos comandos ha sido plasmada en el epígrafe 5.

A.1. Tratamiento de la base de datos original

En este caso, se introducirá el código utilizado para transformar las bases de datos originales (procedentes de *Office for National Statistic*) en sus correspondientes *age- period arrays* (con grupos de periodos y edades que se caracterizan por poseer intervalos de 5 años).

A.1.1. Cifras de suicidios totales

```
# Se elige el directorio donde se encuentra la base de datos original.
# Primero, se debe importar el archivo de excel con los datos:
> library(readxl)
> Base_datos_para_R=read_excel("C:/Users/christian/Desktop/Máster en Estadística Aplicada
UGR/Segundo Semestre/TFM/Archivos R/Base_datos_para_R.xlsx")
# Ya se ha importado la base de datos, ahora se deben agrupar los periodos en intervalos
de 5 años.
# Se crea un dataframe para los totales (eliminando los "rates"). Aquí están los totales.
> Datos_Total=data.frame(Base_datos_para_R)# Primero se crea un data frame vacío.
> borrar_Datos_Total=c("Rate_10_14_years", "Rate_15_19_years", "Rate_20_24_years",
"Rate_25_29_years", "Rate_30_34_years", "Rate_35_39_years", "Rate_40_44_years",
"Rate_45_49_years", "Rate_50_54_years", "Rate_55_59_years", "Rate_60_64_years",
"Rate_65_69_years", "Rate_70_74_years", "Rate_75_79_years", "Rate_80_84_years",
"Rate_85_89_years", "Rate_90_and_over")
> Totales_Numero=Datos_Total[ , !(names(Datos_Total) %in% borrar_Datos_Total)
# Ahora hay que ir separando los totales por sexo:
# Primero hay que quedarse con los totales:
> Totales_Numero_Totales=data.frame(Totales_Numero)
Totales_Numero_Totales=Totales_Numero_Totales[-c(40:117),]# Totales.
> Totales_Numero_Totales=Totales_Numero_Totales[,-1]# Se borra la primera columna
(string).
> Numero_Suicidios_Totales=data.frame(matrix(ncol = 17, nrow = 7))# Nuevo dataframe.
> Numero_Suicidios_Totales[1,]=Totales_Numero_Totales[35,]+Totales_Numero_Totales[34,]
+Totales_Numero_Totales[33,]+Totales_Numero_Totales[32,]+Totales_Numero_Totales[31,]
> Numero_Suicidios_Totales[2,]=Totales_Numero_Totales[30,]+Totales_Numero_Totales[29,]
+Totales_Numero_Totales[28,]+Totales_Numero_Totales[27,]+Totales_Numero_Totales[26,]
> Numero_Suicidios_Totales[3,]=Totales_Numero_Totales[25,]+Totales_Numero_Totales[24,]
```

```

+Totales_Numero_Totales[23,]+Totales_Numero_Totales[22,]+Totales_Numero_Totales[21,]
> Numero_Suicidios_Totales[4,]=Totales_Numero_Totales[20,]+Totales_Numero_Totales[19,]
+Totales_Numero_Totales[18,]+Totales_Numero_Totales[17,]+Totales_Numero_Totales[16,]
> Numero_Suicidios_Totales[5,]=Totales_Numero_Totales[15,]+Totales_Numero_Totales[14,]
+Totales_Numero_Totales[13,]+Totales_Numero_Totales[12,]+Totales_Numero_Totales[11,]
> Numero_Suicidios_Totales[6,]=Totales_Numero_Totales[10,]+Totales_Numero_Totales[9,]
+Totales_Numero_Totales[8,]+Totales_Numero_Totales[7,]+Totales_Numero_Totales[6,]
> Numero_Suicidios_Totales[7,]=Totales_Numero_Totales[5,]+Totales_Numero_Totales[4,]
+Totales_Numero_Totales[3,]+Totales_Numero_Totales[2,]+Totales_Numero_Totales[1,]

# Ahora hay que ponerle nombre a las filas y las columnas:
> colnames(Numero_Suicidios_Totales)=c("Deaths_10_14_years", "Deaths_15_19_years",
"Deaths_20_24_years", "Deaths_25_29_years", "Deaths_30_34_years", "Deaths_35_39_years",
"Deaths_40_44_years", "Deaths_45_49_years", "Deaths_50_54_years", "Deaths_55_59_years",
"Deaths_60_64_years", "Deaths_65_69_years", "Deaths_70_74_years", "Deaths_75_79_years",
"Deaths_80_84_years", "Deaths_85_89_years", "Deaths_90_and_over")
> rownames(Numero_Suicidios_Totales)=c("Period_1985_1989", "Period_1990_1994",
"Period_1995_1999", "Period_2000_2004", "Period_2005_2009", "Period_2010_2014",
"Period_2015_2019")
# Se elimina la fila de los suicidios de personas que tienen entre 10- 14 años:
> Numero_Suicidios_Totales$Deaths_10_14_years=NULL
# También se elimina la columna de personas que tienen 90 o más años:
> Numero_Suicidios_Totales$Deaths_90_and_over=NULL
# Finalmente, queda trasponer la matriz con el objetivo de tener un age- period array
(ahora tenemos un period- age array)
# Se usará la librería data.table:
> library(data.table)
> t_Numero_Suicidios_Totales=transpose(Numero_Suicidios_Totales)
> colnames(t_Numero_Suicidios_Totales)=rownames(Numero_Suicidios_Totales)
> rownames(t_Numero_Suicidios_Totales)=colnames(Numero_Suicidios_Totales)

# Ahora se hará un dataframe para el total de suicidios de varones:
> Totales_Numero_Males=data.frame(Totales_Numero)
# Se eliminan todas las filas que no pertenecen a los totales de hombre (es decir,
las filas de Totales y Totales de mujeres):
> Totales_Numero_Males=Totales_Numero_Males[-c(1:39,79:117),]# Se borra el número de
suicidios totales y el número de suicidios totales por mujeres:
> Totales_Numero_Males=Totales_Numero_Males[,-1]# Se borra la primera columna (que
se correspondía con el período).
> Numero_Suicidios_Males=data.frame(matrix(ncol = 17, nrow = 7))# Se crea un nuevo
dataframe.
> Numero_Suicidios_Males[1,]=Totales_Numero_Males[35,]+Totales_Numero_Males[34,]
+Totales_Numero_Males[33,]+Totales_Numero_Males[32,]+Totales_Numero_Males[31,]
> Numero_Suicidios_Males[2,]=Totales_Numero_Males[30,]+Totales_Numero_Males[29,]
+Totales_Numero_Males[28,]+Totales_Numero_Males[27,]+Totales_Numero_Males[26,]
> Numero_Suicidios_Males[3,]=Totales_Numero_Males[25,]+Totales_Numero_Males[24,]
+Totales_Numero_Males[23,]+Totales_Numero_Males[22,]+Totales_Numero_Males[21,]
> Numero_Suicidios_Males[4,]=Totales_Numero_Males[20,]+Totales_Numero_Males[19,]
+Totales_Numero_Males[18,]+Totales_Numero_Males[17,]+Totales_Numero_Males[16,]
> Numero_Suicidios_Males[5,]=Totales_Numero_Males[15,]+Totales_Numero_Males[14,]
+Totales_Numero_Males[13,]+Totales_Numero_Males[12,]+Totales_Numero_Males[11,]
> Numero_Suicidios_Males[6,]=Totales_Numero_Males[10,]+Totales_Numero_Males[9,]
+Totales_Numero_Males[8,]+Totales_Numero_Males[7,]+Totales_Numero_Males[6,]
> Numero_Suicidios_Males[7,]=Totales_Numero_Males[5,]+Totales_Numero_Males[4,]
+Totales_Numero_Males[3,]+Totales_Numero_Males[2,]+Totales_Numero_Males[1,]
# Ahora hay que poner nombre a las filas y las columnas:

```

```

> colnames(Numero_Suicidios_Males)=c("Deaths_10_14_years", "Deaths_15_19_years",
"Deaths_20_24_years", "Deaths_25_29_years", "Deaths_30_34_years", "Deaths_35_39_years",
"Deaths_40_44_years", "Deaths_45_49_years", "Deaths_50_54_years", "Deaths_55_59_years",
"Deaths_60_64_years", "Deaths_65_69_years", "Deaths_70_74_years", "Deaths_75_79_years",
"Deaths_80_84_years", "Deaths_85_89_years", "Deaths_90_and_over")
> rownames(Numero_Suicidios_Males)=c("Period_1985_1989", "Period_1990_1994",
"Period_1995_1999", "Period_2000_2004", "Period_2005_2009", "Period_2010_2014",
"Period_2015_2019")

# Se elimina la fila que representa los suicidios de las personas que tienen
entre 10 y 14 años:
> Numero_Suicidios_Males$Deaths_10_14_years=NULL
# También se elimina la columna de las personas que tienen 90 o más años:
> Numero_Suicidios_Males$Deaths_90_and_over=NULL
# Ahora falta trasponer la función:
> t_Numero_Suicidios_Males=transpose(Numero_Suicidios_Males)
> colnames(t_Numero_Suicidios_Males)=rownames(Numero_Suicidios_Males)
> rownames(t_Numero_Suicidios_Males)=colnames(Numero_Suicidios_Males)

# Ahora se construirá la base de datos del total de mujeres.
> Totales_Numero_Females=data.frame(Totales_Numero)
# Se eliminan las filas que no pertenecen a los totales de mujeres:
> Totales_Numero_Females=Totales_Numero_Females[-c(1:78),]
> Totales_Numero_Females=Totales_Numero_Females[,-1]# Se borra la primera columna
(periodo):
> Numero_Suicidios_Females=data.frame(matrix(ncol = 17, nrow = 7))# Se crea un nuevo
dataframe:
> Numero_Suicidios_Females[1,]=Totales_Numero_Females[35,]+Totales_Numero_Females[34,]
+Totales_Numero_Females[33,]+Totales_Numero_Females[32,]+Totales_Numero_Females[31,]
> Numero_Suicidios_Females[2,]=Totales_Numero_Females[30,]+Totales_Numero_Females[29,]
+Totales_Numero_Females[28,]+Totales_Numero_Females[27,]+Totales_Numero_Females[26,]
> Numero_Suicidios_Females[3,]=Totales_Numero_Females[25,]+Totales_Numero_Females[24,]
+Totales_Numero_Females[23,]+Totales_Numero_Females[22,]+Totales_Numero_Females[21,]
> Numero_Suicidios_Females[4,]=Totales_Numero_Females[20,]+Totales_Numero_Females[19,]
+Totales_Numero_Females[18,]+Totales_Numero_Females[17,]+Totales_Numero_Females[16,]
> Numero_Suicidios_Females[5,]=Totales_Numero_Females[15,]+Totales_Numero_Females[14,]
+Totales_Numero_Females[13,]+Totales_Numero_Females[12,]+Totales_Numero_Females[11,]
> Numero_Suicidios_Females[6,]=Totales_Numero_Females[10,]+Totales_Numero_Females[9,]
+Totales_Numero_Females[8,]+Totales_Numero_Females[7,]+Totales_Numero_Females[6,]
> Numero_Suicidios_Females[7,]=Totales_Numero_Females[5,]+Totales_Numero_Females[4,]
+Totales_Numero_Females[3,]+Totales_Numero_Females[2,]+Totales_Numero_Females[1,]
# Se pone nombre a las filas y las columnas:
> colnames(Numero_Suicidios_Females)=c("Deaths_10_14_years", "Deaths_15_19_years",
"Deaths_20_24_years", "Deaths_25_29_years", "Deaths_30_34_years", "Deaths_35_39_years",
"Deaths_40_44_years", "Deaths_45_49_years", "Deaths_50_54_years", "Deaths_55_59_years",
"Deaths_60_64_years", "Deaths_65_69_years", "Deaths_70_74_years", "Deaths_75_79_years",
"Deaths_80_84_years", "Deaths_85_89_years", "Deaths_90_and_over")
> rownames(Numero_Suicidios_Females)=c("Period_1985_1989", "Period_1990_1994",
"Period_1995_1999", "Period_2000_2004", "Period_2005_2009", "Period_2010_2014",
"Period_2015_2019")
# Se elimina la fila que representa los suicidios de las mujeres que tienen entre 10
y 14 años:
> Numero_Suicidios_Females$Deaths_10_14_years=NULL
# También se elimina la columna de las mujeres que tienen 90 o más años:
> Numero_Suicidios_Females$Deaths_90_and_over=NULL
# Se traspone la función:

```

```

t_Numero_Suicidios_Females=transpose(Numero_Suicidios_Females)
t_Numero_Suicidios_Females
colnames(t_Numero_Suicidios_Females)=rownames(Numero_Suicidios_Females)
rownames(t_Numero_Suicidios_Females)=colnames(Numero_Suicidios_Females)
# Se pondrán las columnas con nombres parecidos a los que aparecen en el dataset
data.japanese.breast.cancer (dentro de la librería apc):
> rownames(t_Numero_Suicidios_Totales)=c("15-19","20-24","25-29","30-34","35-39",
"40-44","45-49","50-54","55-59", "60-64", "65-69","70-74","75-79","80-84","85-89")
> colnames(t_Numero_Suicidios_Totales)=c("1985-1989","1990-1994","1995-1999",
"2000-2004","2005-2009","2010-2014","2015-2019")
> rownames(t_Numero_Suicidios_Males)=c("15-19","20-24","25-29","30-34","35-39","40-44",
"45-49","50-54","55-59", "60-64", "65-69","70-74","75-79","80-84","85-89")
> colnames(t_Numero_Suicidios_Males)=c("1985-1989","1990-1994","1995-1999","2000-2004",
"2005-2009","2010-2014","2015-2019")
> rownames(t_Numero_Suicidios_Females)=c("15-19","20-24","25-29","30-34","35-39","40-44",
"45-49","50-54","55-59", "60-64", "65-69","70-74","75-79","80-84","85-89")
> colnames(t_Numero_Suicidios_Females)=c("1985-1989","1990-1994","1995-1999","2000-2004",
"2005-2009","2010-2014","2015-2019")

```

A.1.2. Tasas de mortalidad por suicidio

```

> library(readxl)
> Base_datos_Tasas=read_excel("C:/Users/christian/Desktop/Máster en Estadística Aplicada
UGR/Segundo Semestre/TFM/Archivos R/Base_datos_para_R.xlsx")
# Se eliminan las dos primeras columnas de la base de datos (no se trabajará con los
suicidios de personas que tienen entre 10 y 14 años):
> Base_datos_Tasas$Deaths_10_14_years=NULL
> Base_datos_Tasas$Rate_10_14_years=NULL
# Se borran las columnas de las cifras brutas:
> Base_datos_Tasas$Deaths_15_19_years=NULL
> Base_datos_Tasas$Deaths_20_24_years=NULL
> Base_datos_Tasas$Deaths_25_29_years=NULL
> Base_datos_Tasas$Deaths_30_34_years=NULL
> Base_datos_Tasas$Deaths_35_39_years=NULL
> Base_datos_Tasas$Deaths_40_44_years=NULL
> Base_datos_Tasas$Deaths_45_49_years=NULL
> Base_datos_Tasas$Deaths_50_54_years=NULL
> Base_datos_Tasas$Deaths_55_59_years=NULL
> Base_datos_Tasas$Deaths_60_64_years=NULL
> Base_datos_Tasas$Deaths_65_69_years=NULL
> Base_datos_Tasas$Deaths_70_74_years=NULL
> Base_datos_Tasas$Deaths_75_79_years=NULL
> Base_datos_Tasas$Deaths_80_84_years=NULL
> Base_datos_Tasas$Deaths_85_89_years=NULL
> Base_datos_Tasas$Deaths_90_and_over=NULL
# Ahora se tienen, únicamente, los ratios de mortalidad por suicidio (por 100,000 habitantes)
# Hay que ir separando los totales por sexo:
# Primero se trabajará con el total
> Totales_Tasas_Totales=data.frame(Base_datos_Tasas)# Se crea un dataframe
> Totales_Tasas_Totales=Totales_Tasas_Totales[-c(40:117),]# Ahora se tienen los totales
> Totales_Tasas_Totales=Totales_Tasas_Totales[,-c(1)]
> Totales_Tasas_Totales=Totales_Tasas_Totales[,-c(16)]
> Tasas_Suicidios_Totales=data.frame(matrix(ncol = 15, nrow = 7))# Se crea un nuevo
dataframe que se irá completando
> Tasas_Suicidios_Totales[1,]=(Totales_Tasas_Totales[35,]+Totales_Tasas_Totales[34,])

```

```

+Totales_Tasas_Totales[33,]+Totales_Tasas_Totales[32,]+Totales_Tasas_Totales[31,])/5
> Tasas_Suicidios_Totales[2,]=(Totales_Tasas_Totales[30,]+Totales_Tasas_Totales[29,]
+Totales_Tasas_Totales[28,]+Totales_Tasas_Totales[27,]+Totales_Tasas_Totales[26,])/5
> Tasas_Suicidios_Totales[3,]=(Totales_Tasas_Totales[25,]+Totales_Tasas_Totales[24,]
+Totales_Tasas_Totales[23,]+Totales_Tasas_Totales[22,]+Totales_Tasas_Totales[21,])/5
> Tasas_Suicidios_Totales[4,]=(Totales_Tasas_Totales[20,]+Totales_Tasas_Totales[19,]
+Totales_Tasas_Totales[18,]+Totales_Tasas_Totales[17,]+Totales_Tasas_Totales[16,])/5
> Tasas_Suicidios_Totales[5,]=(Totales_Tasas_Totales[15,]+Totales_Tasas_Totales[14,]
+Totales_Tasas_Totales[13,]+Totales_Tasas_Totales[12,]+Totales_Tasas_Totales[11,])/5
> Tasas_Suicidios_Totales[6,]=(Totales_Tasas_Totales[10,]+Totales_Tasas_Totales[9,]
+Totales_Tasas_Totales[8,]+Totales_Tasas_Totales[7,]+Totales_Tasas_Totales[6,])/5
> Tasas_Suicidios_Totales[7,]=(Totales_Tasas_Totales[5,]+Totales_Tasas_Totales[4,]
+Totales_Tasas_Totales[3,]+Totales_Tasas_Totales[2,]+Totales_Tasas_Totales[1,])/5
> colnames(Tasas_Suicidios_Totales)=c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44",
"45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85-89")
> rownames(Tasas_Suicidios_Totales)=c("1985-1989", "1990-1994", "1995-1999", "2000-2004",
"2005-2009", "2010-2014", "2015-2019")

# Se traspone la matriz, para tener un age- period array:
> library(data.table)
> t_Tasas_Suicidios_Totales=transpose(Tasas_Suicidios_Totales)
> t_Tasas_Suicidios_Totales
> colnames(t_Tasas_Suicidios_Totales)=rownames(Tasas_Suicidios_Totales)
> rownames(t_Tasas_Suicidios_Totales)=colnames(Tasas_Suicidios_Totales)
> ls()
> rm(Tasas_Suicidios_Totales,Totales_Tasas_Totales)# Se borran estas bases de datos.
# Ahora se prepara la base de datos para la ratio de mortalidad por suicidio de
varones:
> Totales_Tasas_Males=data.frame(Base_datos_Tasas)# Se crea un dataframe
> Totales_Tasas_Males=Totales_Tasas_Males[-c(1:39,79:117),]
> Totales_Tasas_Males=Totales_Tasas_Males[, -c(1)]
> Totales_Tasas_Males=Totales_Tasas_Males[, -c(16)]# Se borran las personas con 90
o más años.
> Tasas_Suicidios_Males=data.frame(matrix(ncol = 15, nrow = 7))# Se crea un
nuevo dataframe.
> Tasas_Suicidios_Males[1,]=(Totales_Tasas_Males[35,]+Totales_Tasas_Males[34,]
+Totales_Tasas_Males[33,]+Totales_Tasas_Males[32,]+Totales_Tasas_Males[31,])/5
> Tasas_Suicidios_Males[2,]=(Totales_Tasas_Males[30,]+Totales_Tasas_Males[29,]
+Totales_Tasas_Males[28,]+Totales_Tasas_Males[27,]+Totales_Tasas_Males[26,])/5
> Tasas_Suicidios_Males[3,]=(Totales_Tasas_Males[25,]+Totales_Tasas_Males[24,]
+Totales_Tasas_Males[23,]+Totales_Tasas_Males[22,]+Totales_Tasas_Males[21,])/5
> Tasas_Suicidios_Males[4,]=(Totales_Tasas_Males[20,]+Totales_Tasas_Males[19,]
+Totales_Tasas_Males[18,]+Totales_Tasas_Males[17,]+Totales_Tasas_Males[16,])/5
> Tasas_Suicidios_Males[5,]=(Totales_Tasas_Males[15,]+Totales_Tasas_Males[14,]
+Totales_Tasas_Males[13,]+Totales_Tasas_Males[12,]+Totales_Tasas_Males[11,])/5
> Tasas_Suicidios_Males[6,]=(Totales_Tasas_Males[10,]+Totales_Tasas_Males[9,]
+Totales_Tasas_Males[8,]+Totales_Tasas_Males[7,]+Totales_Tasas_Males[6,])/5
> Tasas_Suicidios_Males[7,]=(Totales_Tasas_Males[5,]+Totales_Tasas_Males[4,]
+Totales_Tasas_Males[3,]+Totales_Tasas_Males[2,]+Totales_Tasas_Males[1,])/5
> colnames(Tasas_Suicidios_Males)=c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44",
"45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85-89")
> rownames(Tasas_Suicidios_Males)=c("1985-1989", "1990-1994", "1995-1999", "2000-2004",
"2005-2009", "2010-2014", "2015-2019")
# Se traspone la matriz para tener un age- period array:
> library(data.table)
> t_Tasas_Suicidios_Males=transpose(Tasas_Suicidios_Males)

```

```

> colnames(t_Tasas_Suicidios_Males)=rownames(Tasas_Suicidios_Males)
> rownames(t_Tasas_Suicidios_Males)=colnames(Tasas_Suicidios_Males)
> ls()
> rm(Tasas_Suicidios_Males,Totales_Tasas_Males)# Se borran estas bases de
datos.
# Ahora se tratarán las ratios de mortalidad por suicidio de mujeres.
> Totales_Tasas_Females=data.frame(Base_datos_Tasas)# Se crea un dataframe
> Totales_Tasas_Females=Totales_Tasas_Females[-c(1:78),]
>Totales_Tasas_Females=Totales_Tasas_Females[-c(1)]# Se borra la primera columna.
> Totales_Tasas_Females=Totales_Tasas_Females[-c(16)]# Se borran las personas
con 90 o más años.
> Tasas_Suicidios_Females=data.frame(matrix(ncol = 15, nrow = 7))# Se crea un
dataframe.
> Tasas_Suicidios_Females[1,]=(Totales_Tasas_Females[35,]+Totales_Tasas_Females[34,]
+Totales_Tasas_Females[33,]+Totales_Tasas_Females[32,]+Totales_Tasas_Females[31,])/5
> Tasas_Suicidios_Females[2,]=(Totales_Tasas_Females[30,]+Totales_Tasas_Females[29,]
+Totales_Tasas_Females[28,]+Totales_Tasas_Females[27,]+Totales_Tasas_Females[26,])/5
> Tasas_Suicidios_Females[3,]=(Totales_Tasas_Females[25,]+Totales_Tasas_Females[24,]
+Totales_Tasas_Females[23,]+Totales_Tasas_Females[22,]+Totales_Tasas_Females[21,])/5
> Tasas_Suicidios_Females[4,]=(Totales_Tasas_Females[20,]+Totales_Tasas_Females[19,]
+Totales_Tasas_Females[18,]+Totales_Tasas_Females[17,]+Totales_Tasas_Females[16,])/5
> Tasas_Suicidios_Females[5,]=(Totales_Tasas_Females[15,]+Totales_Tasas_Females[14,]
+Totales_Tasas_Females[13,]+Totales_Tasas_Females[12,]+Totales_Tasas_Females[11,])/5
> Tasas_Suicidios_Females[6,]=(Totales_Tasas_Females[10,]+Totales_Tasas_Females[9,]
+Totales_Tasas_Females[8,]+Totales_Tasas_Females[7,]+Totales_Tasas_Females[6,])/5
> Tasas_Suicidios_Females[7,]=(Totales_Tasas_Females[5,]+Totales_Tasas_Females[4,]
+Totales_Tasas_Females[3,]+Totales_Tasas_Females[2,]+Totales_Tasas_Females[1,])/5
> colnames(Tasas_Suicidios_Females)=c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44",
"45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84", "85-89")
> rownames(Tasas_Suicidios_Females)=c("1985-1989", "1990-1994", "1995-1999", "2000-2004",
"2005-2009", "2010-2014", "2015-2019")
> library(data.table)
> t_Tasas_Suicidios_Females=transpose(Tasas_Suicidios_Females)
> colnames(t_Tasas_Suicidios_Females)=rownames(Tasas_Suicidios_Females)
> rownames(t_Tasas_Suicidios_Females)=colnames(Tasas_Suicidios_Females)
> ls()
> rm(Tasas_Suicidios_Females,Totales_Tasas_Females)# Se borran estas bases de datos
> ls()
> rm(Base_datos_Tasas)
> ls()

```

A.2. Aplicación de los modelos edad- periodo- cohorte mediante el paquete apc

En este epígrafe, se introducirá el código utilizado para implementar los modelos edad- periodo- cohorte aplicados, respectivamente, a las cifras de suicidios y a las tasas de mortalidad por suicidio de Inglaterra y Gales (periodo 1985- 2019).

A.2.1. Cifras de suicidios totales

```

# Se empieza el análisis para las cifras de suicidios totales:
# 1. Construcción del objeto con la instrucción apc.data.list.
> library(apc)
> numero_suicidios_totales_matrix=as.matrix(t_Numero_Suicidios_Totales)

```

```

> Tasas_Suicidios_Totales_matrix=as.matrix(t_Tasas_Suicidios_Totales)
> numero_suicidios_totales=apc.data.list(response = numero_suicidios_totales_matrix,
dose=numero_suicidios_totales_matrix/Tasas_Suicidios_Totales_matrix,
data.format = "AP",age1=15, per1 = 1985, unit = 5,
label = "Número de suicidios totales")
# 2. Análisis gráfico.
> apc.plot.data.all(numero_suicidios_totales)
# 3. Estimación del modelo.
> modelo_numero_suicidios_totales=apc.fit.model(numero_suicidios_totales,
model.family = "poisson.dose.response",model.design = "APC")
> modelo_numero_suicidios_totales
> modelo_numero_suicidios_totales$coefficients.canonical
# Análisis gráfico del modelo estimado:
> apc.plot.fit.all(modelo_numero_suicidios_totales)
# 4. Inferencia sobre el modelo.
> apc.fit.table(numero_suicidios_totales,"poisson.dose.response")

```

A.2.2. Tasas de mortalidad por suicidio

```

# Se analizarán las tasas de mortalidad por suicidio totales:
> library(apc)
# 1. Construcción del objeto con la instrucción apc.data.list.
> tasa_suicidios_totales_matrix=as.matrix(t_Tasas_Suicidios_Totales)
> tasa_suicidios_totales=apc.data.list(tasa_suicidios_totales_matrix,"AP",age1=15,
per1 = 1985, unit = 5, label="Suicide mortality rates England and Wales
(period 1985- 2019)")
> tasa_suicidios_totales
# 2. Análisis gráfico.
> apc.plot.data.all(tasa_suicidios_totales)
# 3. Estimación del modelo usando un APC model.
> modelo_tasa_suicidios_totales=apc.fit.model(tasa_suicidios_totales,
model.family ="gaussian.response",model.design = "APC")
> modelo_tasa_suicidios_totales$coefficients.canonical
# Gráficos del modelo estimado:
> apc.plot.fit.pt(modelo_tasa_suicidios_totales)
> apc.plot.fit(modelo_tasa_suicidios_totales, type = "detrend")
> apc.plot.fit.residuals(modelo_tasa_suicidios_totales)
# 4. Inferencia sobre el modelo.
> deviance_table_apcmodel_total_rates=apc.fit.table(modelo_tasa_suicidios_totales,
"gaussian.response")
> deviance_table_apcmodel_total_rates
#### 5. Estimación del modelo usando un AC model.
> modelo_tasa_suicidios_totales_ACmodel=apc.fit.model(tasa_suicidios_totales,
model.family = "gaussian.response",model.design = "AC")
> modelo_tasa_suicidios_totales_ACmodel$coefficients.canonical
> apc.plot.fit.pt(modelo_tasa_suicidios_totales_ACmodel)
> apc.plot.fit.residuals(modelo_tasa_suicidios_totales_ACmodel)

```

A.2.3. Tasas de mortalidad por suicidio (segmentando por sexos)

```

# Se realiza la comparación entre las tasas de mortalidad por suicidio de mujeres
y hombres
> library(apc)
# 1. Construcción de los objetos con la instrucción apc.data.list.
# Primero se construyen las matrices:
> tasa_suicidios_hombres_matrix=as.matrix(t_Tasas_Suicidios_Males)

```

```

> tasa_suicidios_mujeres_matrix=as.matrix(t_Tasas_Suicidios_Females)
# Construimos los objetos con apc.data.list:
> tasa_suicidios_hombres=apc.data.list(tasa_suicidios_hombres_matrix,"AP",age1=15,
per1 = 1985, unit = 5, label="Suicide mortality rates England and Wales
(period 1985- 2019) for men")
> tasa_suicidios_hombres
> tasa_suicidios_mujeres=apc.data.list(tasa_suicidios_mujeres_matrix,"AP",age1=15,
per1 = 1985, unit = 5, label="Suicide mortality rates England and Wales
(period 1985- 2019) for women")
> tasa_suicidios_mujeres
# 2. Análisis gráfico.
> apc.plot.data.all(tasa_suicidios_hombres)
> apc.plot.data.all(tasa_suicidios_mujeres)
# 3. Estimación de los modelos edad- periodo- cohorte.
> modelo_tasa_suicidios_males=apc.fit.model(tasa_suicidios_hombres,
model.family = "gaussian.response",model.design = "APC")
> modelo_tasa_suicidios_males$coefficients.canonical
> modelo_tasa_suicidios_females=apc.fit.model(tasa_suicidios_mujeres,
model.family = "gaussian.response",model.design = "APC")
> modelo_tasa_suicidios_females$coefficients.canonical
# Gráficos del modelo estimado:
> apc.plot.fit.pt(modelo_tasa_suicidios_males)
> apc.plot.fit.residuals(modelo_tasa_suicidios_males)
> apc.plot.fit.pt(modelo_tasa_suicidios_females)
> apc.plot.fit.residuals(modelo_tasa_suicidios_females)
> apc.plot.fit(modelo_tasa_suicidios_males,type="detrend")
> apc.plot.fit(modelo_tasa_suicidios_females,type="detrend")
# 4. Inferencia sobre el modelo.
> deviance_table_men=apc.fit.table(tasa_suicidios_hombres,"gaussian.response")
> deviance_table_men
> deviance_table_women=apc.fit.table(tasa_suicidios_mujeres,"gaussian.response")
> deviance_table_women

```