



Universidade de Vigo

Trabajo Fin de Máster

Análisis del exceso de mortalidad en Galicia

Ana Domínguez Fernández

Máster en Técnicas Estadísticas

Curso 2023-2024

Propuesta de Trabajo Fin de Máster

Título en galego: Análise do exceso de mortalidade en Galicia
Título en español: Análisis del exceso de mortalidad en Galicia
English title: Analysis of the excess of mortality in Galicia
Modalidad: Modalidad B
Autora: Ana Domínguez Fernández, Universidad de Santiago de Compostela
Directora: Rosa María Crujeiras Casais, Universidad de Santiago de Compostela
Tutora: Gael Naveira Barbeito, Dirección Xeral de Saúde Pública
Breve resumen del trabajo: <p>Este trabajo tiene como objetivo identificar el exceso o defecto de muertes en un año concreto, a partir de las defunciones ocurridas en años pasados. Además, se pretende identificar las causas que contribuyen a las variaciones identificadas. En primer lugar, se realizó una revisión bibliográfica sobre los métodos de análisis existentes para determinar estas diferencias. En segundo lugar, se identificaron las técnicas más adecuadas para su cuantificación. Por último, se aplicaron dichas técnicas a los datos de mortalidad de Galicia durante los últimos cuatro años disponibles empleando como período histórico los diez años previos.</p>

Doña Rosa María Crujeiras Casais, Catedrática de Universidad del área de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela y doña Gael Naveira Barbeito, Técnica Estadística de la Dirección Xeral de Saúde Pública, informan que el Trabajo Fin de Máster titulado

Análisis del exceso de mortalidad en Galicia

fue realizado bajo su dirección por doña Ana Domínguez Fernández para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 3 de junio de 2024.

La directora:
Doña Rosa María Crujeiras Casais

La tutora:
Doña Gael Naveira Barbeito

La autora:
Doña Ana Domínguez Fernández

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, [Disposición 2978 del BOE núm. 48 de 2022](#)), **la autora declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas, . . .)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración, . . . sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Agradecimientos

Me gustaría mostrar mi agradecimiento a mis tutoras, Rosa y Gael, por su orientación y ayuda durante todo el proceso de realización de este trabajo. Y a Soly por su apoyo en el desarrollo del mismo.

Índice general

Resumen	XI
1. Introducción	1
2. Metodología	9
2.1. Modelos lineales generalizados	9
2.1.1. El modelo de regresión de Poisson	11
2.1.2. El modelo de regresión Binomial Negativa	13
2.1.3. Contraste de sobre-dispersión	14
2.2. Introducción a los modelos mixtos	14
2.2.1. Estimación de los parámetros	15
2.2.2. Selección de modelos	18
2.2.3. Validación del modelo	18
2.2.4. Estructura de los modelos del Capítulo 3	19
2.2.5. Estructura de los modelos del Apéndice D	20
3. Aplicación a los datos	23
3.1. Modelos lineales generalizados	25
3.2. Modelos mixtos	29
3.2.1. Año 2019	29
3.2.2. Año 2020	36
3.2.3. Año 2021	43
3.2.4. Año 2022	50
4. Conclusiones y limitaciones	59
A. MoMo y EuroMOMO	61
B. Tasa bruta, específica y ajustada de mortalidad	65
C. Lista reducida de causas de muerte CIE-10 y su correspondencia con la CIE-9	67
D. Otros modelos	71
Bibliografía	101
Índice de Tablas	103
Índice de Figuras	105

Resumen

Resumen en español

Este trabajo se ha desarrollado bajo la modalidad B de los trabajos de fin de máster del Máster en Técnicas Estadísticas, a través de unas prácticas realizadas en el Servicio de Epidemiología de la Dirección Xeral de Saúde Pública (DXSP) de la Consellería de Sanidade. El objetivo principal de esta memoria es identificar el exceso o defecto de muertes en un año concreto, a partir de las defunciones ocurridas en años previos. El objetivo secundario es identificar las causas que contribuyen a las variaciones identificadas.

Este análisis se justificó por un notable aumento en el exceso de muertes en Galicia para el año 2022, detectado por el sistema MoMo (monitorización de la mortalidad diaria por todas las causas), y que la DXSP estudió más en profundidad.

En primer lugar, se llevó a cabo una revisión bibliográfica sobre los métodos de análisis existentes para predecir las defunciones esperadas y así poder determinar la diferencia entre las muertes observadas y las esperadas. En segundo lugar, se realizó la identificación de las técnicas más adecuadas para conseguir la mejor predicción posible, entre estas técnicas se encontraron los modelos lineales generalizados y los modelos mixtos. Por último, se ajustaron los modelos con los datos de mortalidad de Galicia de los diez años anteriores al año a predecir, obteniendo así las defunciones esperadas para cada uno de los años del período 2019-2022.

English abstract

This work has been developed under modality B of the final projects for the Master's in Statistical Techniques, through an internship carried out at the Epidemiology Service of the Dirección Xeral de Saúde Pública (DXSP) of the Consellería de Sanidade. The main objective of this report is to identify the excess or deficit of deaths in a specific year, based on deaths that occurred in previous years. The secondary objective is to identify the causes that contribute to the observed variations.

This analysis was justified by a notable increase in excess deaths in Galicia for the year 2022, detected by the MoMo system (daily mortality monitoring for all causes), which prompted a more in-depth study by the DXSP.

First, a bibliographic review was conducted on the existing analysis methods for predicting expected deaths, to determine the difference between observed and expected deaths. Second, the most appropriate techniques for achieving the best possible prediction were identified, including generalized linear models and mixed models. Finally, the models were fitted using the mortality data from Galicia from the ten years preceding the year to be predicted, thus obtaining the expected deaths for each of the years in the period 2019-2022.

Capítulo 1

Introducción

Esta memoria se ha desarrollado bajo la modalidad B de los trabajos fin de máster del Máster en Técnicas Estadísticas, a través de unas prácticas realizadas en el Servicio de Epidemiología de la Dirección Xeral de Saúde Pública (DXSP) de la Consellería de Sanidade, con el objetivo de hacer un estudio sobre el exceso o defecto de mortalidad en Galicia.

A nivel España se cuenta con el sistema de Monitorización de la Mortalidad (MoMo) diaria por todas las causas. Este sistema, coordinado por el Ministerio de Sanidad, se ha implementado en la Unidad de Vigilancia de la Mortalidad diaria del Centro Nacional de Epidemiología (CNE) del Instituto de Salud Carlos III. Uno de sus objetivos es identificar las desviaciones de la mortalidad diaria observada con respecto a la esperada según las series históricas de mortalidad, lo que permite estimar de forma indirecta el impacto de cualquier evento de importancia en Salud Pública. En el Panel MoMo se actualizan diariamente las estimaciones de excesos de mortalidad por todas las causas y atribuibles a un exceso o defecto de la temperatura, por ámbito poblacional, sexo y grupo de edad. Los datos están disponibles en https://momo.isciii.es/panel_momo/. Para obtener información adicional acerca de MoMo y EuroMOMO, que es un proyecto europeo de vigilancia de la mortalidad, se puede consultar el [Apéndice A](#).

A partir de los resultados obtenidos de MoMo para Galicia, recogidos en la [Tabla 1.1](#), se identificaron excesos anuales en torno al 4-5 % sobre el total de defunciones anuales observadas, desde el año 2016 hasta el 2021, a excepción del año 2019 para el que se estimó un defecto de muertes. En el año 2022 hubo un exceso de 3350 defunciones, lo que supuso un 9.5 % del total de muertes observadas ese año.

Año	Defunciones observadas	Defunciones estimadas	Exceso por todas las causas	Porcentaje de exceso
2015	32098	29701	2397	7.47 %
2016	32210	30445	1765	5.48 %
2017	32450	31289	1161	3.58 %
2018	32886	31435	1452	4.42 %
2019	31721	31867	-146	-0.46 %
2020	33042	31626	1416	4.29 %
2021	32956	31254	1702	5.16 %
2022	35164	31814	3350	9.53 %

Tabla 1.1: Defunciones anuales observadas, estimadas, exceso por todas las causas y porcentaje de excesos anuales sobre el total de defunciones anuales observadas. Galicia, 2015-2022. Fuente: MoMo.

Este aumento, identificado por MoMo, en el exceso de muertes del año 2022 justificó un estudio más en profundidad, realizado por la DXSP, del comportamiento de la mortalidad en Galicia.

Los datos de mortalidad se obtuvieron del Registro de Mortalidad de Galicia (Dirección Xeral de Saúde Pública) y la población a partir de la Renovación Anual del Padrón (Instituto Galego de Estatística¹). En primer lugar, se estudió la evolución de las tasas anuales brutas y ajustadas (ver definiciones en el [Apéndice B](#)), tomando como población estándar el censo de 2011. En la [Figura 1.1](#) se puede observar cómo la tasa bruta de mortalidad en Galicia presenta una tendencia ascendente. Sin embargo, al ajustar por edad esta tendencia se invierte, lo que va en consonancia con el progresivo envejecimiento de la población. Con todo, y para las dos tasas, se aprecia un descenso en el año 2019, un comportamiento estable para 2020-2021 y un aumento en 2022.

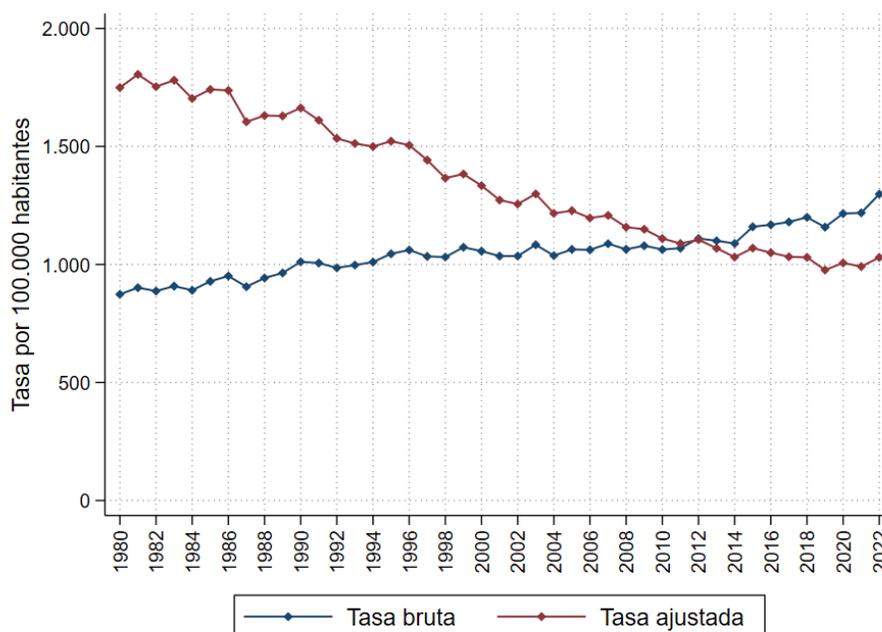


Figura 1.1: Evolución de la tasa bruta de mortalidad y la tasa ajustada por edad (población estándar: censo del 2011) en Galicia en el período 1980-2022.

En segundo lugar, la DXSP analizó el comportamiento de la mortalidad en los años 2020, 2021 y 2022. Para ello, se estimaron las tasas específicas de mortalidad (ver definición en el [Apéndice B](#)) por grupos de edad quinquenales, a partir de la mortalidad observada en el período 2015-2019. Estas tasas se aplicaron a las poblaciones de los tres años siguientes para estimar el número de defunciones esperadas en esos años, bajo la hipótesis de que se mantuviese el patrón de mortalidad del período 2015-2019.

En la [Figura 1.2](#) se aprecia que desde el año 2015 al 2018 la mortalidad observada presentó una tendencia lineal con una pendiente creciente pero no muy pronunciada. Al llegar al año 2019 se identificó un descenso importante de la mortalidad, pasando de 32416 a 31184 defunciones. En el año 2020 se recuperó la mortalidad y en el 2021 se mantuvo constante el número de defunciones con respecto al año previo, ambos sucesos probablemente relacionados con las medidas de protección adoptadas frente al COVID-19. Por último, en el 2022 se observó un aumento acusado del número de defunciones pasando de 32853 en el 2021 a 34707 en el 2022.

Por otro lado, la mortalidad esperada en los años 2020-2022 mantuvo una tendencia similar a lo observado en el período 2015-2018. Entonces, comparando los datos de la mortalidad observada y esperada para esos tres años, se identificó un defecto de defunciones en el año 2020, más acusado en el 2021. En el año 2022 parece recuperarse la tendencia prepandemia, coincidiendo prácticamente en su

¹Instituto Galego de Estatística: <https://www.ige.gal>

totalidad las defunciones observadas con las esperadas. El descenso de la mortalidad que se produjo en el 2019 y el estancamiento de las muertes en el 2021 es información que emplea MoMo para estimar las defunciones esperadas en el 2022. Por lo tanto, estos dos sucesos podrían estar explicando el aumento en el exceso de muertes del año 2022, detectado por MoMo.

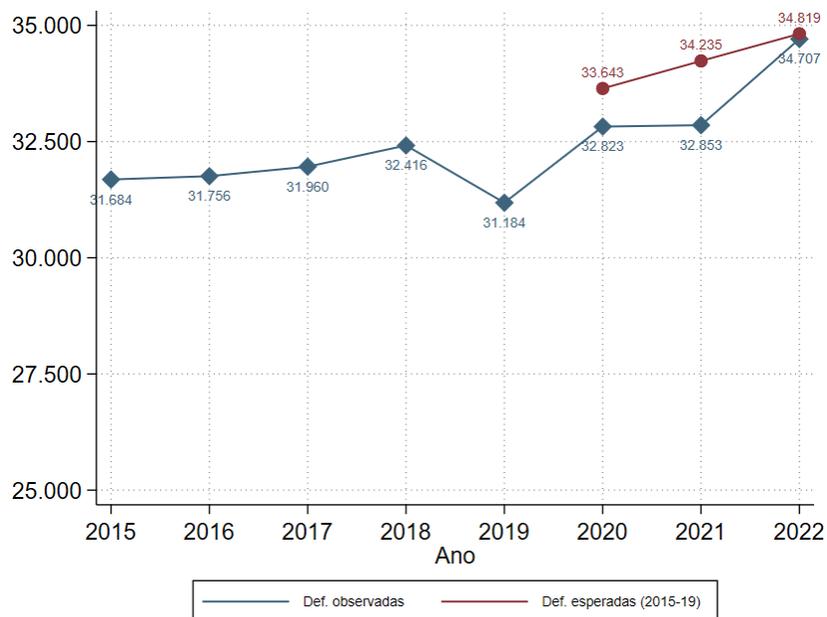


Figura 1.2: Evolución de la mortalidad observada en Galicia desde el año 2015 al 2022 y de la mortalidad esperada desde el año 2020 al 2022. En el eje *X* están representados los años y en el eje *Y* el número de defunciones.

A continuación, empleando la misma metodología, se calcularon las defunciones esperadas en función de la causa de muerte. La lista de capítulos de enfermedades, recogidos en la 10ª revisión de la Clasificación Internacional de Enfermedades (CIE-10) que se puede consultar en el [Apéndice C](#), es la siguiente:

- Capítulo *I*: Enfermedades infecciosas y parasitarias
- Capítulo *II*: Tumores
- Capítulo *III*: Enfermedades de la sangre/órganos hematopoiéticos/inmunidad
- Capítulo *IV*: Enfermedades endocrinas/nutricionales/metabólicas
- Capítulo *V*: Trastornos mentales y de comportamiento
- Capítulo *VI*: Enfermedades del sistema nervioso
- Capítulo *VII*: Enfermedades del ojo y sus anexos
- Capítulo *VIII*: Enfermedades del oído y de la apófisis mastoides
- Capítulo *IX*: Enfermedades del sistema circulatorio
- Capítulo *X*: Enfermedades del sistema respiratorio
- Capítulo *XI*: Enfermedades del sistema digestivo

- Capítulo *XII*: Enfermedades de la piel y del tejido subcutáneo
- Capítulo *XIII*: Enfermedades del sistema osteomuscular y tejido conectivo
- Capítulo *XIV*: Enfermedades del sistema genitourinario
- Capítulo *XV*: Embarazo, parto y puerperio
- Capítulo *XVI*: Afecciones originadas en el período perinatal
- Capítulo *XVII*: Malformaciones congénitas/anomalías cromosómicas
- Capítulo *XVIII*: Síntomas/signos/resultados anormales, no clasificados bajo otro concepto
- Capítulo *XX*: Causas externas de mortalidad

En función del peso y relevancia que supuso cada uno de los capítulos en la mortalidad global fueron seleccionados para estudiar individualmente los capítulos *I*, *II*, *IX*, *X*, *XIII*, *XVIII* y *XX*. El resto de capítulos fueron considerados conjuntamente y se denominó “Resto”.

En la [Figura 1.3](#) se observa que el capítulo *I* presentó un exceso de defunciones en los tres años; en cambio los capítulos *II*, *IX* y *X* presentaron un defecto de muertes. El capítulo *XVIII*, que es el de causas mal definidas, aumentó en gran medida en el 2022 respecto de lo que sería esperable.

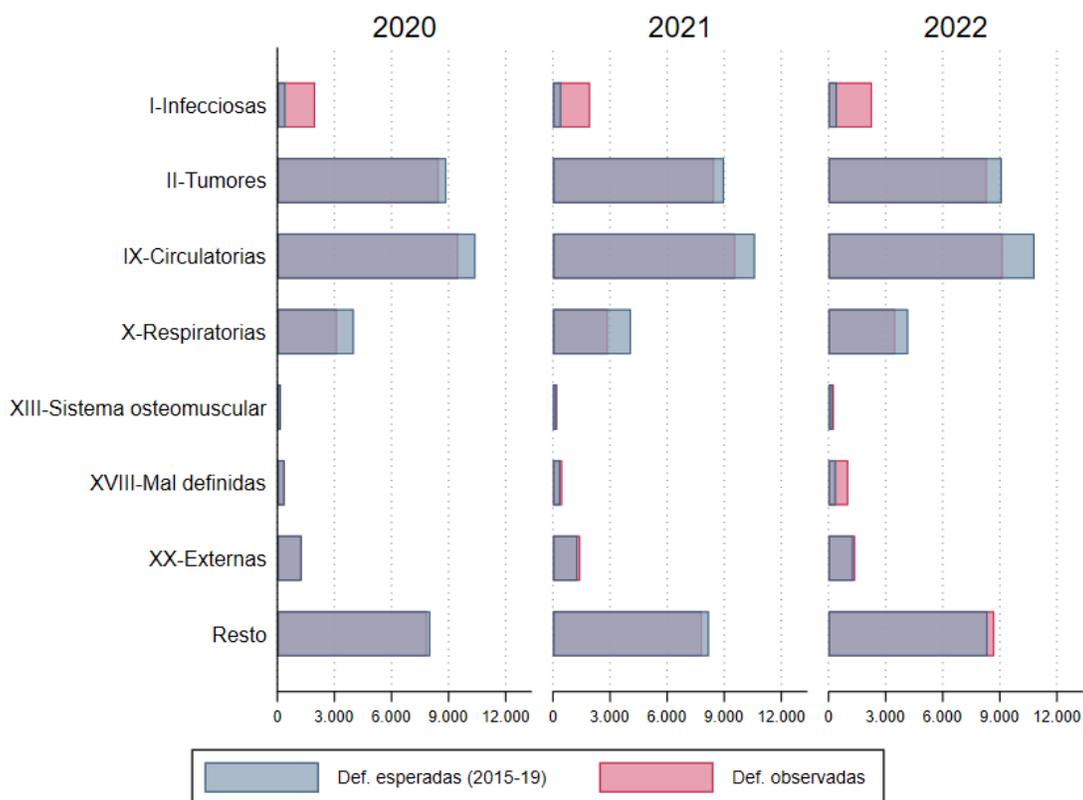


Figura 1.3: Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función del capítulo de la causa de muerte.

Una posible justificación de estos excesos y defectos sería que muchas de las personas que fallecieron en exceso por enfermedades infecciosas serían posibles candidatos a fallecer por alguna enfermedad del sistema circulatorio o respiratorio, y de ahí los defectos detectados en esos capítulos. En particular, en el 2022 se observó un exceso de 1847 muertes en el capítulo *I*; mientras que en los capítulos *IX* y *X* se observó un defecto de 1669 y 680 muertes, respectivamente.

En el caso de los tumores, una posible justificación apunta a la falta de información o la no revisión de historias clínicas que pueden conllevar a que una muerte por cáncer acabe siendo enmascarada por otra causa de defunción, por lo general, más inespecífica.

El exceso de muertes en los capítulos *I* y *XVIII* fue lo que justificó un estudio detallado en lista reducida para el cálculo de las defunciones esperadas. La lista reducida es una subclasificación más específica de causas de muerte dentro de los capítulos (ver [Apéndice C](#)).

Estos dos capítulos fueron desglosados en causas de la lista reducida. Como era de esperar en el capítulo *I*, enfermedades infecciosas y parasitarias, se observa en la [Figura 1.4](#) un exceso de mortalidad por COVID-19, tanto confirmado como sospechoso, superior a 1500 fallecidos en los tres años. Esto se debe a que en el período del 2015 al 2019 estas causas no existían, por ello la mortalidad esperada era nula. Esta situación justifica en casi su totalidad el exceso observado en el capítulo *I*. El resto de enfermedades infecciosas también presentaron un exceso de muertes en el año 2022, entorno a 100 personas más.

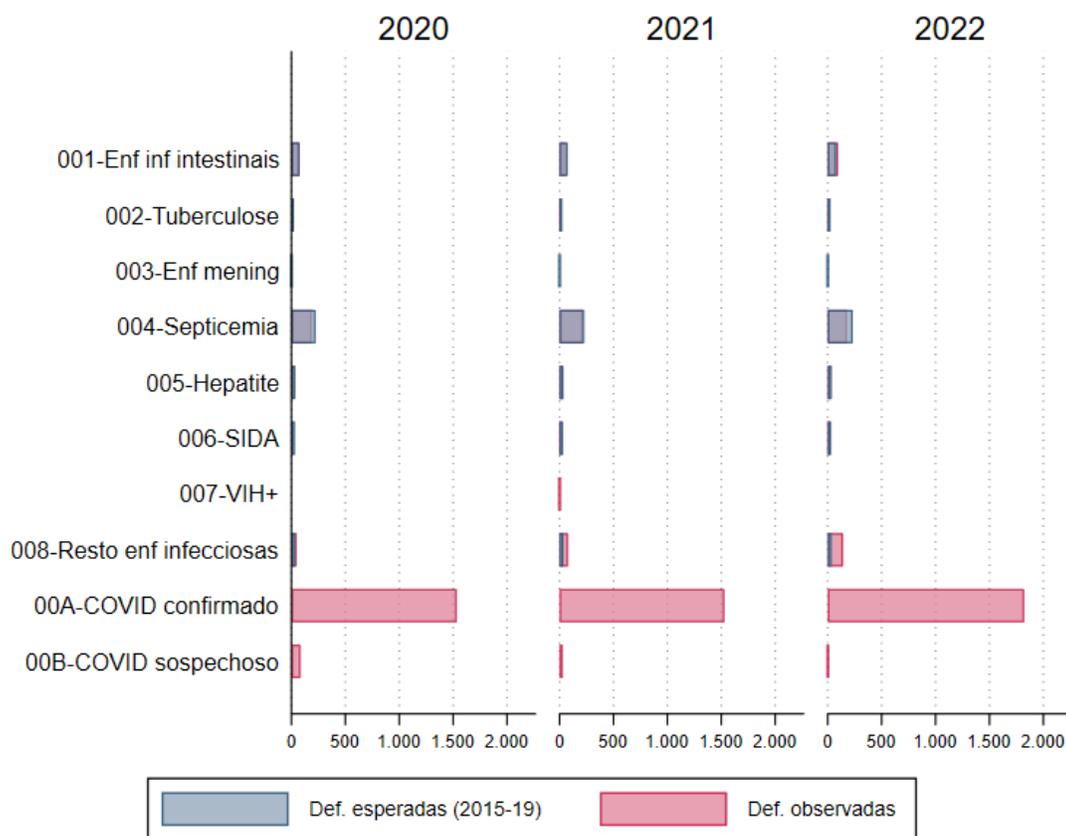


Figura 1.4: Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función de la lista reducida de causas de muerte del capítulo *I*.

Por otro lado, se tiene el capítulo *XVIII* que es el que incluye las causas consideradas mal definidas, representadas por términos médicos que no aportan información desde el punto de vista clínico y epidemiológico para establecer estrategias de prevención y control de las enfermedades. Estas causas se analizan para estudiar errores en la información sobre la causa de muerte. La calidad de los datos sobre las causas de muerte depende de la fiabilidad de la certificación de la defunción, realizada por un profesional médico, así como de la precisión en la codificación de la causa de muerte, llevada a cabo por un codificador que no es el médico que certificó la muerte.

En la [Figura 1.5](#) se observa que la senilidad y resto de síntomas presentaron un exceso de mortalidad en los tres años, aunque más pronunciado en el año 2022. El paro cardíaco mostró un defecto, en comparación con lo que cabría esperar, en los años 2020 y 2021 con 103 y 106 muertes de menos, respectivamente, y pasó a tener un exceso en el 2022 con 99 muertes de más. En global, para todo el capítulo, los excesos de muerte fueron en aumento en estos tres años, llegando a 642 muertes más que las esperadas en el año 2022. Esta situación, relativa principalmente a 2022, así como el defecto de muertes en el capítulo *II* de tumores, se enmarca temporalmente dentro de un proceso de cambio en el personal codificador del Registro de Mortalidad de Galicia, que conllevó la renovación total de la plantilla y la correspondiente formación de la misma.

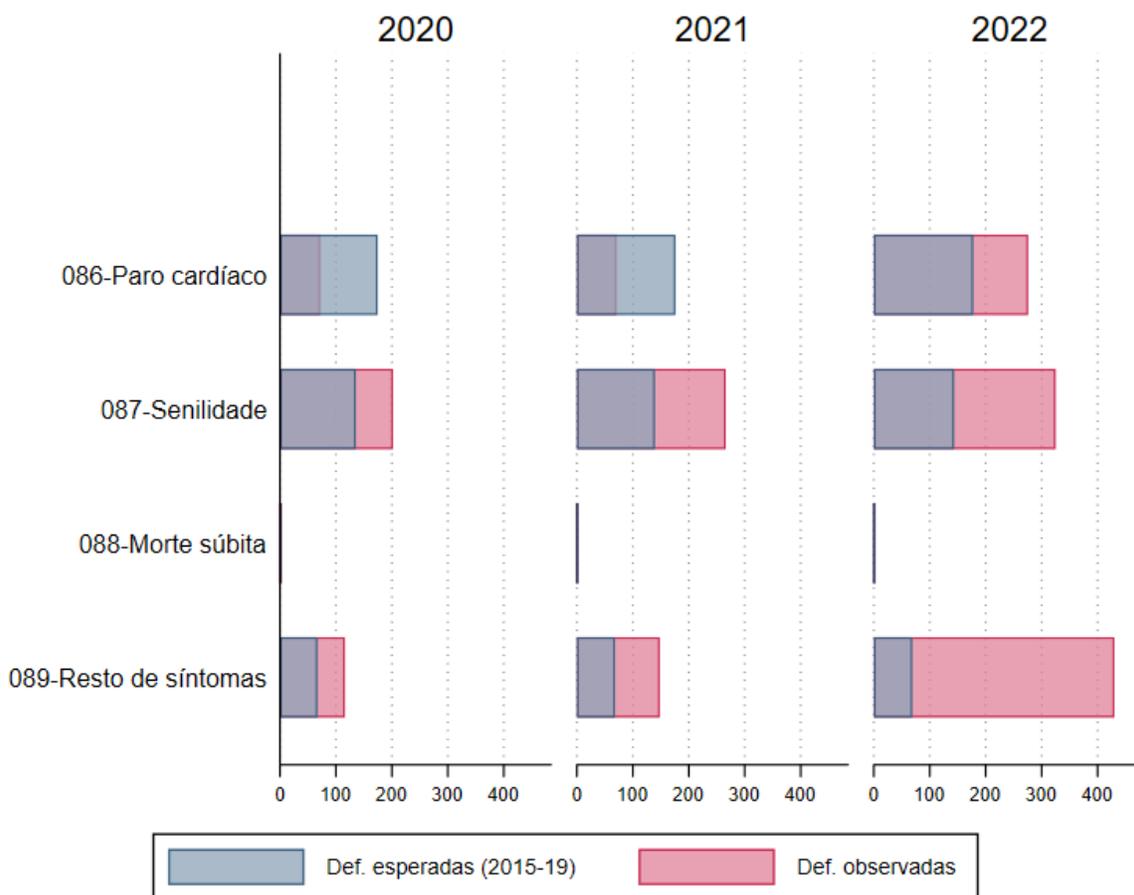


Figura 1.5: Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función de la lista reducida de causas de muerte del capítulo *XVIII*.

A parte del capítulo *XVIII*, existen otros códigos que conllevan una mala codificación y se encuentran en capítulos distintos, por lo que se hizo un estudio centrándose en una lista más completa

de códigos mal definidos. Fuera del capítulo *XVIII*, las afecciones consideradas mal definidas son las siguientes:

- Muerte cardíaca repentina (código: *I46.1* del capítulo *IX*)
- Paro cardíaco no especificado (código: *I46.9* del capítulo *IX*)
- Fallo cardíaco agudo (código: *I50.9* del capítulo *IX*)
- Hipotensión, no especificada (código: *I95.9* del capítulo *IX*)
- Enfermedad circulatoria, no especificada (código *I99* del capítulo *IX*)
- Insuficiencia respiratoria aguda (código: *J96.0* del capítulo *X*)
- Insuficiencia respiratoria, no especificada (código: *J96.9* del capítulo *X*)
- Insuficiencia respiratoria de recién nacidos (código: *P28.5* del capítulo *XVI*)

Al tener en cuenta el conjunto de enfermedades consideradas mal definidas, se observa en la [Figura 1.6](#) que hubo un exceso de muertes en el año 2021 con 219 muertes de más y de forma más acusada en el año 2022 con 802 muertes de más.

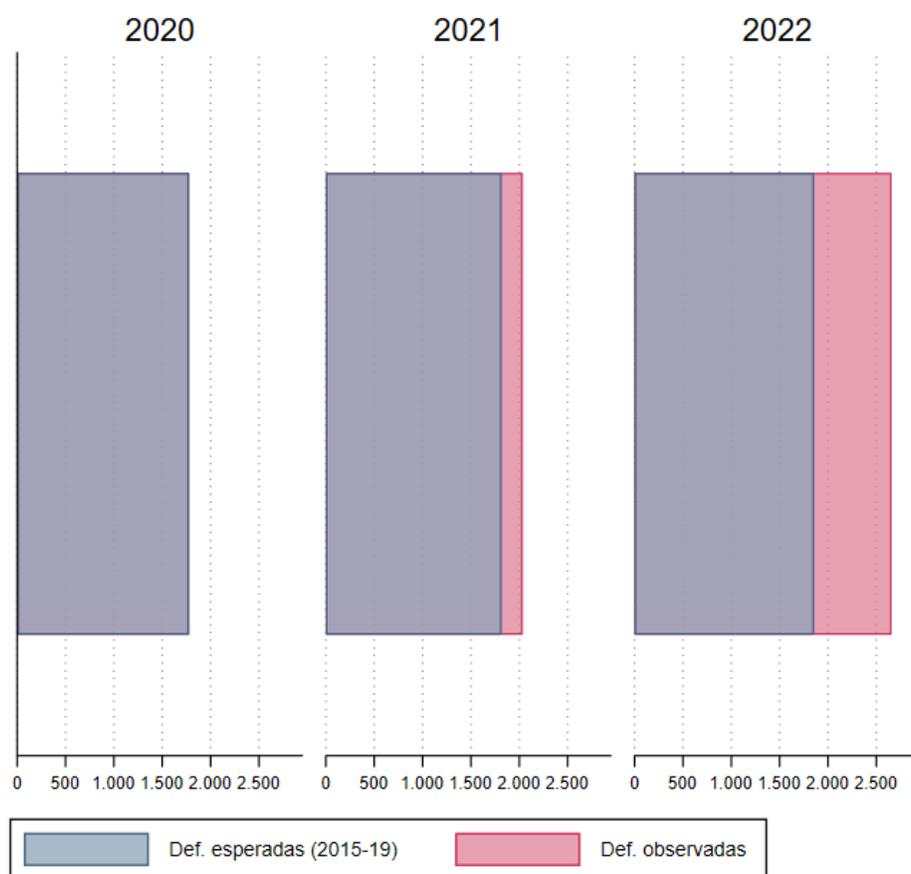


Figura 1.6: Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 para el conjunto de las causas mal definidas.

Por lo tanto, la motivación y el objetivo de este trabajo son:

- **Motivación.** A partir de este análisis, desde la DXSP se identificó la necesidad de establecer una metodología que permita, de forma general, identificar cambios en los patrones de mortalidad en un determinado año, por exceso o por defecto, y explicar esos cambios en términos de causas de muerte.
- **Objetivo.** El objetivo de este trabajo es ajustar modelos de predicción de las defunciones esperadas para un determinado año, con respecto a un período histórico inmediatamente anterior, teniendo en cuenta la causa de muerte en el modelo. La comparación de las defunciones esperadas con las observadas permitirá establecer la existencia de un exceso o un defecto de muertes e identificar las causas asociadas.

A continuación se detalla la estructura de esta memoria. En el [Capítulo 2](#), se introducen y estudian en detalle los modelos que se han encontrado en la revisión bibliográfica sobre la estimación del exceso de defunciones. En particular, se estudian el modelo de regresión de Poisson, el modelo de regresión Binomial Negativa y los modelos mixtos. En el [Capítulo 3](#), se emplean los modelos estudiados en el [Capítulo 2](#) para ajustar los datos de interés con el fin de ver cuál de los métodos es el más adecuado para el ajuste. En el [Capítulo 4](#), están recogidas las conclusiones de los resultados obtenidos con los ajustes realizados en el [Capítulo 3](#), así como las limitaciones de los modelos ajustados. En el [Apéndice A](#), se encuentra información adicional acerca de MoMo y EuroMOMO. En el [Apéndice B](#), se proporcionan las definiciones de tasa bruta, específica y ajustada de mortalidad. En el [Apéndice C](#), se puede consultar la lista reducida de causas de muerte CIE-10 y su correspondencia con la CIE-9. En el [Apéndice D](#), se aplica a los datos uno de los modelos mixtos explicados en el [Capítulo 3](#).

Para llevar a cabo los distintos análisis realizados se utilizó Stata y . En particular, la función `nbreg` de Stata se empleó para ajustar el modelo de regresión Binomial Negativa, mientras que en  se utilizó el paquete `glmmTMB` para ajustar los modelos mixtos. Los datos proporcionados por la DXSP, al ser provenientes del Registro de Mortalidad de Galicia, no necesitaron una depuración. A partir de ellos, se estableció la estructura y agregación necesaria para poder llevar a cabo el estudio.

Capítulo 2

Metodología

En este Capítulo, se realiza un estudio de la metodología existente para estimar las defunciones esperadas, con la finalidad de identificar cuál de los métodos disponibles es el más apropiado para hacer la estimación. Para realizar este estudio metodológico, se llevó a cabo una búsqueda bibliográfica, lo que implicó revisar trabajos previos relacionados con la predicción de defunciones. Además, se realizó un curso sobre búsquedas bibliográficas para asegurar una revisión exhaustiva y precisa de la literatura existente. Como resultado, se identificaron varios enfoques diferentes, que se han agrupado en cuatro bloques principales: los modelos lineales generalizados, las series de tiempo por si solas o combinadas con los modelos anteriores, los modelos jerárquicos bayesianos y los modelos funcionales.

En la búsqueda bibliográfica realizada, los artículos que se basan en métodos de series de tiempo son [Foreman et al. \(2018\)](#), [Achebak et al. \(2019\)](#) y [Vieira et al. \(2020\)](#). En contraste, [Santos et al. \(2021\)](#), [Islam et al. \(2021\)](#) y [Shin et al. \(2021\)](#) utilizan modelos lineales generalizados, en particular el modelo de Poisson con sobre-dispersión, el modelo binomial negativo y la regresión cuasi-Poisson, respectivamente, para abordar la estimación de muertes esperadas. En este bloque también se encuentra el artículo [Tawiah et al. \(2021\)](#) donde se usan distintos modelos lineales generalizados y se comparan para ver cual obtiene mejores resultados. Por otro lado, [Blangiardo et al. \(2020\)](#) opta por un modelo jerárquico bayesiano, mientras que en [Verbeeck et al. \(2023\)](#) se emplea el modelo lineal mixto.

A continuación, en las distintas secciones del Capítulo se introducen brevemente cada uno de los modelos que han sido utilizados en el contexto de interés.

2.1. Modelos lineales generalizados

Para encontrar una introducción completa a este tipo de modelos se puede consultar el capítulo 3 de [Agresti \(2007\)](#). En un modelo lineal generalizado (GLM, del inglés Generalized Linear Model) se considera una variable respuesta Y y un vector de variables explicativas $X' = (X_1, \dots, X_d)$. Se supone una distribución de probabilidad de la variable respuesta (conocida) y cuyos parámetros, que dependen de las variables explicativas, se pretenden estimar. Se denota cada una de las observaciones de Y por Y_i con $i = 1, \dots, n$ donde n es el tamaño muestral de Y .

Por ejemplo, si cada Y_i es el número de éxitos en una secuencia de ensayos con respuesta dicotómica se tiene que Y sigue una distribución binomial. Si cada Y_i es un recuento, es decir, toma números enteros no negativos, se puede asumir como distribución de Y la Poisson o la Binomial Negativa. Si cada observación es continua se puede suponer una distribución continua para Y como por ejemplo, la normal, la Gamma... en función de las características de la variable Y .

La formulación general de un GLM es la siguiente:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$$

donde $\mu = \mathbb{E}(Y|X = \mathbf{x})$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_d)'$ es el vector de parámetros $(d + 1)$ -dimensional y $\mathbf{x} = (1, x_1, \dots, x_d)'$ es el vector $(d + 1)$ -dimensional que contiene las variables explicativas y el término de

intercepto. La combinación lineal de las variables explicativas con el vector de parámetros se denomina predictor lineal y $g(\cdot)$ es la función link o función de enlace que es la que permite relacionar μ con el predictor lineal.

Cada observación de la variable respuesta de los datos que se van a analizar en el [Capítulo 3](#) es un recuento, por tanto, el estudio de los GLMs se centra en el modelo de Poisson y el de la Binomial Negativa.

Antes de empezar con el estudio de ambos modelos, se introduce la familia exponencial de distribuciones. A esta familia pertenecen la distribución Normal, la Binomial, la de Poisson, la Binomial Negativa, etc. La función de densidad de la familia exponencial tiene la siguiente expresión

$$f_Y(y; \theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \quad (2.1)$$

y su log-verosimilitud es

$$l(y; \theta, \phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

donde las funciones a , b y c son conocidas, θ es el parámetro canónico y ϕ es un parámetro de dispersión.

Las tres funciones y los dos parámetros cambian en función de la distribución que se quiera emplear. En particular, se van a obtener estas tres funciones para la distribución de Poisson y para la Binomial Negativa. En primer lugar, con la distribución de Poisson se tiene que $Y \sim \text{Poisson}(\mu)$ entonces su función de masa de probabilidad viene dada por

$$P(Y = k) = \frac{e^{-\mu} \mu^k}{k!}, \quad k = \{0, 1, \dots\}$$

que puede reescribirse como:

$$P(Y = k) = \exp [k \log(\mu) - \mu - \log(k!)], \quad (2.2)$$

se toma $\theta = \log(\mu)$ entonces $\mu = e^\theta$, sustituyendo estas dos igualdades en (2.2) se obtiene la siguiente igualdad:

$$P(Y = k) = \exp [k\theta - e^\theta - \log(k!)]. \quad (2.3)$$

La igualdad (2.3) ahora presenta una estructura equivalente a la de (2.1). En consecuencia, se puede concluir que

$$b(\theta) = e^\theta, \quad a(\phi) = 1 \quad \text{y} \quad c(y, \phi) = -\log(k!).$$

Con la distribución Binomial Negativa se tiene que si $Y \sim \text{BN}(k; r, p)$ entonces su función de masa de probabilidad viene dada por

$$\begin{aligned} P(Y = k) &= \binom{k+r-1}{k} (1-p)^r p^k = \frac{(k+r-1)!}{k!(r-1)!} (1-p)^r p^k \\ &\stackrel{(a)}{=} \frac{\Gamma(k+r)}{k! \Gamma(r)} (1-p)^r p^k, \quad k = \{0, 1, \dots\} \end{aligned} \quad (2.4)$$

donde $\Gamma(\cdot)$ es la función gamma y en la igualdad (a) se aplica la siguiente propiedad: $\Gamma(n+1) = n!$. Entonces, (2.4) puede reescribirse como:

$$\begin{aligned} P(Y = k) &= \binom{k+r-1}{k} (1-p)^r p^k \\ &= \exp \left[k \log(p) + r \log(1-p) + \log \binom{k+r-1}{k} \right], \end{aligned} \quad (2.5)$$

se toma $\theta = \log(p)$ entonces $p = e^\theta$, sustituyendo estas dos igualdades en (2.5) se obtiene la siguiente igualdad:

$$P(Y = k) = \exp \left[k\theta + r \log(1 - e^\theta) + \log \binom{k+r-1}{k} \right]. \quad (2.6)$$

La igualdad (2.6) ahora presenta una estructura equivalente a la de (2.1). En consecuencia, se puede concluir que

$$b(\theta) = -r \log(1 - e^\theta), \quad a(\phi) = 1 \quad \text{y} \quad c(y, \phi) = \log \binom{k+r-1}{k}.$$

Dos propiedades importantes de la familia exponencial que se usarán más adelante son:

$$\mathbb{E}(Y) = \mu = \frac{\partial b(\theta)}{\partial \theta} \quad \text{y} \quad (2.7)$$

$$\text{Var}(Y) = a(\phi) \left[\frac{\partial^2 b(\theta)}{\partial \theta^2} \right]. \quad (2.8)$$

Se puede ver que estas dos propiedades se cumplen para la distribución Binomial Negativa, teniendo en cuenta que $b(\theta) = -r \log(1 - e^\theta)$ y $a(\phi) = 1$, se obtienen los siguientes resultados para el cálculo de la media y la varianza:

$$\mathbb{E}(Y) = \frac{\partial [-r \log(1 - e^\theta)]}{\partial \theta} = \frac{r e^\theta}{1 - e^\theta} \stackrel{p=e^\theta}{=} \frac{pr}{1 - p} \quad \text{y}$$

$$\text{Var}(Y) = \frac{\partial^2 [-r \log(1 - e^\theta)]}{\partial \theta^2} = \frac{r e^\theta (1 - e^\theta) + r e^{2\theta}}{(1 - e^\theta)^2} = \frac{r e^\theta}{(1 - e^\theta)^2} \stackrel{p=e^\theta}{=} \frac{pr}{(1 - p)^2}.$$

Con la familia exponencial explicada, se procede al estudio del modelo de regresión Poisson y del modelo de regresión Binomial Negativa.

2.1.1. El modelo de regresión de Poisson

Cuando los valores de la variable respuesta son números enteros no negativos y se corresponden con un proceso de conteo, es habitual considerar una distribución de Poisson (véase el capítulo 9 de [Zuur et al. \(2009\)](#)). Por esta razón, se necesita una función link para no tener problemas con los números negativos a la hora de estimar el modelo lineal. Una posible opción para la función link es tomar $g(u) = \log(u)$, $u \in (0, \infty)$ cuya inversa es $g^{-1}(v) = e^v$, $v \in \mathbb{R}$.

La función de regresión es la media de Y condicionada a cada valor de la explicativa, es decir,

$$m(\mathbf{x}) = \mu(\mathbf{x}) = \mathbb{E}(Y|X = \mathbf{x}).$$

Usando el logaritmo como función link se obtiene que

$$g(\mu(\mathbf{x}, \boldsymbol{\beta})) = \mathbf{x}'\boldsymbol{\beta}. \quad (2.9)$$

Aplicando la función inversa del logaritmo a (2.9) se tiene que la función de regresión del modelo de Poisson se expresa como:

$$\mu(\mathbf{x}, \boldsymbol{\beta}) = e^{\mathbf{x}'\boldsymbol{\beta}}. \quad (2.10)$$

Estimación e inferencia

Como $Y_i \sim \text{Poisson}(\mu(X_i, \boldsymbol{\beta}))$ se sabe que la función de masa de probabilidad condicional para cada $Y_i = y_i$ viene dada por

$$f(y_i) = e^{-\mu(x_i, \boldsymbol{\beta})} \frac{\mu(x_i, \boldsymbol{\beta})^{y_i}}{y_i!},$$

por tanto la función de verosimilitud es la siguiente

$$f(y_1, \dots, y_n) = L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[e^{-\mu(x_i, \boldsymbol{\beta})} \frac{\mu(x_i, \boldsymbol{\beta})^{y_i}}{y_i!} \right].$$

Se aplica el logaritmo a la función de verosimilitud obteniendo que la log-verosimilitud viene dada por

$$\begin{aligned} \log(L(\boldsymbol{\beta})) = l(\boldsymbol{\beta}) &= \sum_{i=1}^n \left[\log \left(e^{-\mu(x_i, \boldsymbol{\beta})} \frac{\mu(x_i, \boldsymbol{\beta})^{y_i}}{y_i!} \right) \right] \\ &= \sum_{i=1}^n \left[-\mu(x_i, \boldsymbol{\beta}) + \log \left(\frac{\mu(x_i, \boldsymbol{\beta})^{y_i}}{y_i!} \right) \right] \\ &= \sum_{i=1}^n \left[-\mu(x_i, \boldsymbol{\beta}) + y_i \log(\mu(x_i, \boldsymbol{\beta})) - \log(y_i!) \right], \end{aligned}$$

utilizando la expresión (2.10) se sustituye $\mu(x_i, \boldsymbol{\beta})$ obteniendo la siguiente expresión

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \left[-e^{x_i' \boldsymbol{\beta}} + y_i \log(e^{x_i' \boldsymbol{\beta}}) - \log(y_i!) \right] = \sum_{i=1}^n \left[-e^{x_i' \boldsymbol{\beta}} + x_i' y_i \boldsymbol{\beta} - \log(y_i!) \right],$$

la última fórmula se deriva respecto a $\boldsymbol{\beta}$ y se iguala a cero

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left(-x_i' e^{x_i' \boldsymbol{\beta}} + x_i' y_i \right) = \sum_{i=1}^n x_i' \left(y_i - e^{x_i' \boldsymbol{\beta}} \right) = 0.$$

Estas ecuaciones no tienen una solución explícita entonces para obtener las estimaciones se necesitan procedimientos iterativos, en este caso, los más populares son el método de Newton-Raphson y el IRWLS (iteratively reweighted least squares) que pueden consultarse en [Faraway \(2016\)](#).

Para construir intervalos de confianza y realizar contrastes de hipótesis sobre los parámetros se puede recurrir a tres métodos. En primer lugar, se tiene la distribución asintótica teniendo en cuenta que el estimador es asintóticamente insesgado y la matriz de covarianzas es la matriz hessiana de $l(\boldsymbol{\beta})$ que tiene la siguiente expresión

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} = - \sum_{i=1}^n x_i x_i' e^{x_i' \boldsymbol{\beta}} = - \sum_{i=1}^n x_i x_i' \mu(x_i, \boldsymbol{\beta}).$$

En segundo lugar, se puede emplear el perfil de verosimilitud para la construcción de intervalos de confianza. Y por último, para llevar a cabo un contraste de un modelo simplificado con uno más general se puede utilizar la deviance. ¹

¹Estos tres procedimientos se estudiaron en la materia de Regresión Generalizada y Modelos Mixtos del Máster en Técnicas Estadísticas de la Universidad de Santiago de Compostela.

Validación del modelo

Para la validación del modelo se estudian los residuos obtenidos a partir de la estimación de los parámetros:

$$y_i - \hat{\mu}(x_i, \hat{\beta}).$$

La regresión de Poisson permite tener una mayor variación de estos residuos que un modelo lineal con suposición de normalidad en los errores, entonces esto hace que no sean particularmente útiles. Dos tipos de residuos que se pueden usar en su lugar son:

- Los residuos de Pearson que se obtienen al dividir los residuos por la desviación típica de Y_i :

$$\frac{y_i - \hat{\mu}(x_i, \hat{\beta})}{\sqrt{\widehat{\text{Var}}(y_i)}} = \frac{y_i - \hat{\mu}(x_i, \hat{\beta})}{\sqrt{\hat{\mu}(x_i, \hat{\beta})}}$$

- Los residuos de la deviance tienen la siguiente expresión

$$\text{signo} \left(y_i - \hat{\mu}(x_i, \hat{\beta}) \right) \sqrt{2 \left[y_i \log \left(y_i / \hat{\mu}(x_i, \hat{\beta}) \right) - \left(y_i - \hat{\mu}(x_i, \hat{\beta}) \right) \right]}$$

2.1.2. El modelo de regresión Binomial Negativa

Si una variable sigue una distribución de Poisson se está asumiendo que la media y la varianza son iguales, pero hay situaciones en las que los datos de recuento suelen variar más de lo que cabría esperar si la distribución de la respuesta realmente fuera Poisson. En particular, si la varianza es mayor que la media se está dando lo que se conoce como fenómeno de sobre-dispersión.

Si se da este fenómeno de sobre-dispersión, se puede ajustar la regresión Binomial Negativa (véase el capítulo 9 de [Zuur et al. \(2009\)](#)) que, como en la Poisson, la variable respuesta toma valores en $\{0, 1, 2, 3, \dots\}$. Sea $Y \sim \text{BN}(k; r, p)$ entonces su función de masa de probabilidad viene dada por la expresión (2.4). Para escribir la función de distribución en función de la media, μ , y del parámetro de sobre-dispersión, r , que es un parámetro no negativo se toma

$$\mu = \frac{pr}{1-p},$$

cuanto mayor sea el valor del parámetro de sobre-dispersión, menor será la sobre-dispersión. A partir de la igualdad de μ se despeja p ,

$$\mu = \frac{pr}{1-p} \implies \mu - \mu p = pr \implies \mu = \mu p + pr \implies \mu = p(\mu + r) \implies p = \frac{\mu}{r + \mu}$$

de aquí se obtienen la esperanza y la varianza de la variable X en función de μ y r ,

$$\begin{aligned} \mathbb{E}(Y) &= \frac{pr}{1-p} = \mu \quad y \\ \text{Var}(Y) &= \frac{pr}{(1-p)^2} = \mu \frac{1}{\left(1 - \frac{\mu}{r+\mu}\right)} = \mu \left(\frac{r+\mu}{r} \right) = \mu + \frac{\mu^2}{r}. \end{aligned}$$

Por lo tanto, si se supone que la variable respuesta sigue una distribución Binomial Negativa, es decir, que $Y_i \sim \text{BN}(\mu(X_i, \beta), r)$ se tiene que

$$\mathbb{E}(Y_i) = \mu(X_i, \beta) \quad y \quad \text{Var}(Y_i) = \mu(X_i, \beta) + \frac{\mu(X_i, \beta)^2}{r}.$$

De igual manera que se hizo con el modelo de Poisson se usa como función link el logaritmo, llegando a que la función de regresión del modelo de regresión Binomial Negativa se expresa como:

$$\mu(x, \beta) = e^{x'\beta}.$$

2.1.3. Contraste de sobre-dispersión

El contraste de sobre-dispersión es equivalente a hacer un test de razón de verosimilitudes entre el modelo de Poisson y el modelo de regresión Binomial Negativa. La función de masa de probabilidad de la Binomial Negativa para $k = \{0, 1, \dots\}$, en función del r y de μ , tiene la siguiente expresión:

$$P(Y = k) = \frac{\Gamma(k+r)}{k!\Gamma(r)} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^k = \frac{\Gamma(k+r)}{\Gamma(r)(r+\mu)^k} \frac{\mu^k}{k!} \frac{1}{\left(1+\frac{\mu}{r}\right)^r}. \quad (2.11)$$

Si se toma el límite cuando r tiende a infinito se llega al siguiente resultado:

$$\lim_{r \rightarrow \infty} P(Y = k) = \frac{e^{-\mu} \mu^k}{k!}, \quad k = \{0, 1, \dots\}$$

que viene de que cuando $r \rightarrow \infty$: el primer término de la segunda igualdad de (2.11) tiende a 1, el segundo no depende de r por tanto se considera un término constante y en el tercero se emplea que el denominador tiende a e^μ . Es decir, se ha llegado a que la masa de probabilidad de la Binomial Negativa tiende a la masa de probabilidad de la Poisson cuando $r \rightarrow \infty$.

Por lo tanto, se plantea el siguiente contraste de hipótesis sobre el parámetro de sobre-dispersión:

$$\begin{cases} H_0 : r = \infty, \mu > 0 \\ H_1 : r < \infty, \mu > 0 \end{cases} \quad (2.12)$$

Con el contraste (2.12) se comprueba si hay o no sobre-dispersión. Si el contraste no encuentra evidencias contra la hipótesis nula esto querría decir que no se puede asumir sobre-dispersión (modelo de regresión de Poisson), en cambio si rechaza H_0 quiere decir que se puede asumir sobre-dispersión (modelo de regresión Binomial Negativa).

Sea $f_Y(y_1, \dots, y_n; r, \mu)$ la verosimilitud de la Binomial Negativa, entonces se define la razón de verosimilitud del test como:

$$\Lambda(y_1, \dots, y_n) = \frac{\sup_{(r, \mu) \in \Theta_0} f_Y(y_1, \dots, y_n; r, \mu)}{\sup_{(r, \mu) \in \Theta_0 \cup \Theta_1} f_Y(y_1, \dots, y_n; r, \mu)} = \frac{f_Y(y_1, \dots, y_n; \hat{r}_0, \hat{\mu}_0)}{f_Y(y_1, \dots, y_n; \hat{r}, \hat{\mu})}$$

donde \hat{r}_0 y $\hat{\mu}_0$ son los estimadores de máxima verosimilitud de f considerando Θ_0 y \hat{r} y $\hat{\mu}$ son los estimadores de máxima verosimilitud de f considerando $\Theta_0 \cup \Theta_1$. Siendo Θ_0 el espacio paramétrico de H_0 y $\Theta_0 \cup \Theta_1$ el espacio paramétrico al que pertenece r , es decir, $\mathbb{N} \cup \{0\}$. La expresión explícita de estos estimadores puede consultarse en Villaseñor Alva et al. (2001).

Por último, para el estadístico de contraste se puede consultar el Teorema 9.1 en Vélez Ibarrola y García Pérez (2012) que bajo ciertas condiciones de regularidad se cumple que

$$-2 \log(\Lambda(y_1, \dots, y_n)) \xrightarrow{d} \chi_1^2$$

donde χ_1^2 es la distribución χ^2 con un grado de libertad. Los grados de libertad de la distribución son la diferencia entre los parámetros que se estiman en H_1 menos los parámetros que se estiman en H_0 . En este caso, en H_1 se estima un parámetro más que en H_0 (el parámetro de sobre-dispersión r), por este motivo la distribución χ^2 tiene un grado de libertad.

2.2. Introducción a los modelos mixtos

Para encontrar una introducción completa a este tipo de modelos puede consultarse Stroup (2012). Los modelos mixtos son modelos con efectos fijos y efectos aleatorios que se emplean principalmente

para estudiar relaciones entre una variable respuesta y algunas variables explicativas. Son especialmente útiles cuando se quiere incorporar diversas fuentes de variabilidad en el modelo, incorporando no solo errores a nivel individual, como se observa, por ejemplo, en datos agrupados. Los datos longitudinales, las medidas repetidas o los datos multinivel son casos particulares de datos agrupados. En este trabajo, los datos que se emplearán en el [Capítulo 3](#) para ajustarlos con los modelos explicados en este Capítulo, son datos longitudinales ya que la variable de interés se mide a lo largo del tiempo.

La formulación general de los modelos mixtos es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (2.13)$$

donde \mathbf{Y} es el vector respuesta n -dimensional, \mathbf{X} es la matriz de diseño con dimensión $n \times d$ con d el número de efectos fijos, $\boldsymbol{\beta}$ es el vector de parámetros de efectos fijos d -dimensional, \mathbf{Z} es una matriz con dimensión $n \times q$ con q el número de efectos aleatorios, \mathbf{b} es el vector de efectos aleatorios q -dimensional y $\boldsymbol{\varepsilon}$ es el vector de errores n -dimensional. Es decir, la parte fija del modelo está representada por $\mathbf{X}\boldsymbol{\beta}$ y la parte aleatoria por $\mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$.

En general, se supone que $\mathbf{Y}|\mathbf{b}$ sigue una distribución con $\mathbb{E}(\mathbf{Y}|\mathbf{b}) = \boldsymbol{\mu}|\mathbf{b}$ y $\text{Var}(\mathbf{Y}|\mathbf{b}) = \mathbf{R}$, y que $\boldsymbol{\varepsilon}$ sigue la misma distribución con un vector de medias $\mathbf{0}$ n -dimensional y una matriz de varianzas-covarianzas \mathbf{R} de dimensión $n \times n$. Se define $\boldsymbol{\eta}$ como el predictor lineal $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$ donde $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$ y existe una función link g tal que $\boldsymbol{\eta} = g(\boldsymbol{\mu}|\mathbf{b})$. Sea $h(\cdot) = g^{-1}(\cdot)$, entonces aplicando la función inversa de g se tiene que

$$\boldsymbol{\mu} = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}).$$

2.2.1. Estimación de los parámetros

A continuación, se supone que el modelo solo tiene efectos fijos, es decir, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. Para la estimación de los parámetros de este modelo se aplica una modificación del método de máxima verosimilitud. Teniendo en cuenta que $\boldsymbol{\theta}$ es una función de la media, es decir, $\boldsymbol{\theta}(\boldsymbol{\mu})$ y que $\boldsymbol{\mu} = h(\boldsymbol{\eta})$ se llega a que $\boldsymbol{\theta} = \boldsymbol{\theta}[h(\mathbf{X}\boldsymbol{\beta})]$. Por tanto, para obtener el estimador de máxima verosimilitud de los efectos fijos, $\boldsymbol{\beta}$, se maximiza la log-verosimilitud de la familia exponencial respecto a $\boldsymbol{\beta}$. La expresión matricial de la log-verosimilitud es

$$\begin{aligned} l(\boldsymbol{\theta}; \mathbf{Y}; \boldsymbol{\phi}) &= \mathbf{Y}'\mathbf{A}\boldsymbol{\theta} - \mathbf{1}'\mathbf{A}\mathbf{b}(\boldsymbol{\theta}) + c(\mathbf{Y}, \boldsymbol{\phi}) \\ &= \mathbf{Y}'\mathbf{A} \{ \boldsymbol{\theta}[h(\mathbf{X}\boldsymbol{\beta})] \} - \mathbf{1}'\mathbf{A}\mathbf{b}(\{ \boldsymbol{\theta}[h(\mathbf{X}\boldsymbol{\beta})] \}) + c(\mathbf{Y}, \boldsymbol{\phi}) \end{aligned}$$

donde \mathbf{Y} es el vector respuesta n -dimensional, \mathbf{A} es una matriz diagonal de dimensión $n \times n$ donde el i -ésimo elemento de la diagonal es el inverso del valor del parámetro escala para la observación Y_i , es decir, $\mathbf{A} = \text{diag}[1/a(\phi_i)]$, $\mathbf{1}'$ es el vector fila n -dimensional, $\mathbf{b}(\boldsymbol{\theta})$ es el vector n -dimensional que a la elemento de $\boldsymbol{\theta}$ le aplica la función b y por último, $c(\mathbf{Y}, \boldsymbol{\phi})$ es el término que no depende de $\boldsymbol{\theta}$.

Por simplicidad, se evalúa la derivada escalar de la log-verosimilitud de $f_Y(\boldsymbol{\theta}; y, \boldsymbol{\phi})$ con respecto a $\boldsymbol{\beta}$. En otras palabras, se calcula la siguiente derivada:

$$\frac{\partial l(\boldsymbol{\theta}[h(\mathbf{X}\boldsymbol{\beta})]; y, \boldsymbol{\phi})}{\partial \boldsymbol{\beta}} = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}} = \frac{y - \boldsymbol{\mu}}{a(\boldsymbol{\phi})} \left(\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right)^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \right) \mathbf{X}, \quad (2.14)$$

donde en la segunda igualdad se han usado los siguientes cálculos:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{y - \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{a(\boldsymbol{\phi})} \stackrel{(a)}{=} \frac{y - \boldsymbol{\mu}}{a(\boldsymbol{\phi})}, \quad \text{en (a) se ha usado la igualdad (2.7),} \\ \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}} &= \left(\frac{\partial \left[\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]}{\partial \boldsymbol{\theta}} \right)^{-1} = \left(\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right)^{-1} \quad y \end{aligned}$$

$$\frac{\partial \eta}{\partial \beta} = \frac{\partial(\mathbf{X}\beta)}{\partial \beta} = \mathbf{X}.$$

Por último, aplicando la propiedad (2.8) de la familia exponencial a (2.14) se obtiene que

$$\frac{\partial l(\theta[h(\mathbf{X}\beta)]; y, \phi)}{\partial \beta} = (y - \mu) \left(\frac{1}{V(y)} \right) \left(\frac{\partial \mu}{\partial \eta} \right) \mathbf{X},$$

cuya expresión matricial es

$$\frac{\partial l(\boldsymbol{\theta}[h(\mathbf{X}\beta)])}{\partial \beta'} = \mathbf{X}'\mathbf{D}^{-1}\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \quad (2.15)$$

donde $l(\boldsymbol{\theta})$ es el vector n -dimensional donde el i -ésimo elemento es la log-verosimilitud de cada observación Y_i , \mathbf{D} es la matriz diagonal de dimensión $n \times n$ donde los elementos de la diagonal son las derivadas $\partial \eta_i / \partial \mu_i$ y \mathbf{V} es la matriz diagonal de dimensión $n \times n$ donde el i -ésimo elemento de la diagonal es la varianza de Y_i .

Se define la matriz \mathbf{W} como sigue $\mathbf{W} = (\mathbf{D}\mathbf{V}\mathbf{D})^{-1}$, entonces ahora se puede reescribir la expresión (2.15) como

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}[h(\mathbf{X}\beta)])}{\partial \beta'} &= \mathbf{X}'\mathbf{D}^{-1}\mathbf{V}^{-1}(\mathbf{D}^{-1}\mathbf{D})(\mathbf{Y} - \boldsymbol{\mu}) \\ &= \mathbf{X}'(\mathbf{D}\mathbf{V}\mathbf{D})^{-1}\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{X}'\mathbf{W}\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu}) \end{aligned}$$

Para llegar a las ecuaciones de la estimación, se usa la ecuación del algoritmo de Fisher scoring². Este algoritmo se puede consultar en la Sección 4.2.3 de [Stroup \(2012\)](#). Antes de aplicar la ecuación mencionada, es necesario calcular la varianza de $\frac{\partial l(\boldsymbol{\theta}[h(\mathbf{X}\beta)])}{\partial \beta'}$:

$$\begin{aligned} \text{Var} \left(\frac{\partial l(\boldsymbol{\theta}[h(\mathbf{X}\beta)])}{\partial \beta'} \right) &= \text{Var}(\mathbf{X}'\mathbf{W}\mathbf{D}(\mathbf{Y} - \boldsymbol{\mu})) = \mathbf{X}'\mathbf{W}\mathbf{D}[\text{Var}(\mathbf{Y} - \boldsymbol{\mu})]\mathbf{D}\mathbf{W}\mathbf{X} \\ &= \mathbf{X}'\mathbf{W}\mathbf{D}\mathbf{V}\mathbf{D}\mathbf{W}\mathbf{X} = \mathbf{X}'\mathbf{W}\mathbf{W}^{-1}\mathbf{W}\mathbf{X} = \mathbf{X}'\mathbf{W}\mathbf{X}. \end{aligned}$$

Con todos los elementos necesarios ya calculados, se aplica la ecuación del algoritmo de Fisher scoring, obteniendo el siguiente resultado

$$\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}} + [\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X}]^{-1} \mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{D}}(\mathbf{Y} - \tilde{\boldsymbol{\mu}})$$

donde $\tilde{\mathbf{W}}$, $\tilde{\mathbf{D}}$ y $\tilde{\boldsymbol{\mu}}$ representan a \mathbf{W} , \mathbf{D} y $\boldsymbol{\mu}$ evaluadas en $\tilde{\boldsymbol{\eta}} = \mathbf{X}\tilde{\boldsymbol{\beta}}$. Ahora, multiplicando en ambos lados de la igualdad por $\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X}$ se llega a que

$$\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\tilde{\mathbf{W}}\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{D}}(\mathbf{Y} - \tilde{\boldsymbol{\mu}}) = \mathbf{X}'\tilde{\mathbf{W}}\mathbf{Y}^* \quad (2.16)$$

donde

$$\mathbf{Y}^* = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{D}}(\mathbf{Y} - \tilde{\boldsymbol{\mu}}) = \tilde{\boldsymbol{\eta}} + \tilde{\mathbf{D}}(\mathbf{Y} - \tilde{\boldsymbol{\mu}}). \quad (2.17)$$

²Sea $\boldsymbol{\theta}$ es el parámetro que se quiere estimar y $\tilde{\boldsymbol{\theta}}$ su estimador, la ecuación del algoritmo de Fisher scoring tiene la siguiente expresión

$$\boldsymbol{\theta} \cong \tilde{\boldsymbol{\theta}} + [\mathbf{I}(\tilde{\boldsymbol{\theta}})]^{-1} \left. \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

donde $\mathbf{I}(\tilde{\boldsymbol{\theta}})$ es la matriz de información de Fisher evaluada en $\tilde{\boldsymbol{\theta}}$. Además, se sabe que $\mathbf{I}(\tilde{\boldsymbol{\theta}}) = \text{Var} \left(\left. \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \right)$

A partir de la ecuación (2.16), se lleva a cabo un proceso iterativo para calcular la estimación del vector de parámetros β .

A continuación, se tienen en cuenta los efectos aleatorios, es decir, $\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{b}$. Además, se cumplen las siguientes hipótesis: $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$, $\mathbb{E}(\mathbf{y}|\mathbf{b}) = \boldsymbol{\mu}|\mathbf{b}$ y $\text{Var}(\mathbf{y}|\mathbf{b}) = \mathbf{V}_\mu^{\frac{1}{2}}\mathbf{A}\mathbf{V}_\mu^{\frac{1}{2}}$ donde $\mathbf{V}_\mu^{\frac{1}{2}} = \text{diag} \left[\sqrt{\partial^2 b(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2} \right]$. En este caso, se emplea la cuasi-verosimilitud marginal para llevar a cabo la estimación de los parámetros β y \mathbf{b} . Para el cálculo de la cuasi-log-verosimilitud se tiene que:

- la cuasi-log-verosimilitud condicionada a los efectos aleatorios es:

$$\text{ql}(\mathbf{Y}|\mathbf{b}) = \mathbf{Y}'\mathbf{A}\boldsymbol{\theta} - \mathbf{1}'\mathbf{A}\mathbf{b}(\boldsymbol{\theta})$$

- y la log-verosimilitud de los efectos aleatorios es:

$$l(\mathbf{b}) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{G}|) - \frac{1}{2} \mathbf{b}'\mathbf{G}^{-1}\mathbf{b}.$$

Por tanto, la fórmula de la cuasi-log-verosimilitud es $\text{ql}(\mathbf{Y}|\mathbf{b}) + l(\mathbf{b})$. A partir de esta expresión se llega a que la cuasi-log-verosimilitud marginal se expresa como sigue:

$$\begin{aligned} \text{ql}(\mathbf{Y}) &= \iint [\text{ql}(\mathbf{Y}|\mathbf{b}) + l(\mathbf{b})] d\mathbf{b} \\ &= \iint \left[\mathbf{Y}'\mathbf{A}\boldsymbol{\theta} - \mathbf{1}'\mathbf{A}\mathbf{b}(\boldsymbol{\theta}) - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{G}|) - \frac{1}{2} \mathbf{b}'\mathbf{G}^{-1}\mathbf{b} \right] d\mathbf{b} \end{aligned}$$

Para aproximar esta integral se usará la pseudo-verosimilitud. En primer lugar, se aplica un desarrollo de Taylor a la función h evaluada en $\tilde{\boldsymbol{\eta}}$, obteniendo así la siguiente aproximación

$$h(\boldsymbol{\eta}) \cong h(\tilde{\boldsymbol{\eta}}) + \left. \frac{\partial h(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}} (\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}). \quad (2.18)$$

En (2.15) se había definido $\mathbf{D} = \text{diag}[\partial \boldsymbol{\eta}/\partial \boldsymbol{\mu}] = \text{diag}[\partial g(\boldsymbol{\mu})/\partial \boldsymbol{\mu}]$ por lo tanto la inversa de esta matriz es $\mathbf{D}^{-1} = \text{diag}[\partial \boldsymbol{\mu}/\partial \boldsymbol{\eta}] = \text{diag}[\partial h(\boldsymbol{\eta})/\partial \boldsymbol{\eta}]$. Con la matriz \mathbf{D}^{-1} ya definida y teniendo en cuenta que $\boldsymbol{\eta} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b}$, se puede reescribir (2.18) como sigue

$$h(\boldsymbol{\eta}) \cong h(\tilde{\boldsymbol{\eta}}) + \tilde{\mathbf{D}}^{-1} (\mathbf{X}\beta + \mathbf{Z}\mathbf{b} - \mathbf{X}\tilde{\beta} - \mathbf{Z}\tilde{\mathbf{b}}),$$

reordenando los términos y multiplicando en ambos de la igualdad por $\tilde{\mathbf{D}}$ se a llega que

$$\tilde{\mathbf{D}} [h(\boldsymbol{\eta}) - h(\tilde{\boldsymbol{\eta}})] + \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{b}} \cong \mathbf{X}\beta + \mathbf{Z}\mathbf{b}.$$

Se definió \mathbf{Y}^* en (2.17), en este caso se define como

$$\mathbf{Y}^* = \tilde{\boldsymbol{\eta}} + \tilde{\mathbf{D}} [\mathbf{Y} - \tilde{\boldsymbol{\mu}}|\tilde{\mathbf{b}}] = \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{b}} + \tilde{\mathbf{D}} [\mathbf{Y} - h(\tilde{\boldsymbol{\eta}})],$$

se calcula ahora la esperanza y varianza condicionada de \mathbf{Y}^* :

$$\begin{aligned} \mathbb{E}(\mathbf{Y}^*|\mathbf{b}) &= \tilde{\mathbf{D}} [h(\boldsymbol{\eta}) - h(\tilde{\boldsymbol{\eta}})] + \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\mathbf{b}} \cong \mathbf{X}\beta + \mathbf{Z}\mathbf{b} \\ \text{Var}(\mathbf{Y}^*|\mathbf{b}) &= \mathbf{D}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{A}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{D}. \end{aligned}$$

donde la inversa de $\mathbf{D}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{A}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{D}$ se denota por \mathbf{W} , es decir, $\mathbf{W} = (\mathbf{D}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{A}\mathbf{V}_\mu^{\frac{1}{2}}\mathbf{D})^{-1}$. Con esto, se obtienen las siguientes ecuaciones de pseudo-verosimilitud para estimar los vectores de parámetros $\boldsymbol{\beta}$ y \mathbf{b} :

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{X}'\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{Y}^* \\ \mathbf{Z}'\mathbf{W}\mathbf{Y}^* \end{bmatrix}$$

Los cálculos de como llegar a estas ecuaciones se pueden consultar en la Sección 4.5 de [Stroup \(2012\)](#). También se pueden consultar otros métodos a parte de la pseudo-verosimilitud, como pueden ser la aproximación de Laplace o la cuadratura de Gauss-Hermite (ambos son métodos de aproximación integral).

2.2.2. Selección de modelos

Para la selección de modelos se pueden emplear el criterio de información de Akaike (AIC) y el criterio de información Bayesiano (BIC). Ambos criterios introducen un término de penalización para el número de parámetros del modelo, el término de penalización es mayor en el BIC que en el AIC.

- La fórmula del criterio de información de Akaike es

$$\text{AIC} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) + 2s$$

donde s es el número total de parámetros a estimar en el modelo y $l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}})$ es la función de log-verosimilitud evaluada en las estimaciones de $\boldsymbol{\beta}$ y \mathbf{b} .

- La fórmula del criterio de información Bayesiano es

$$\text{BIC} = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) + s \log(n)$$

donde n es el tamaño muestral.

Dados dos modelos estimados, el modelo con el menor valor de BIC es el que se prefiere. Y ocurre lo mismo con el AIC.

2.2.3. Validación del modelo

Debido a la presencia de los efectos aleatorios, \mathbf{b} , y las particularidades en los métodos de estimación utilizados, la validación de estos modelos se vuelve más compleja en comparación a la de los modelos lineales generalizados. En este contexto, la varianza de los residuos en el primer nivel no necesariamente debe ser constante, por lo tanto se pueden calcular los residuos de la deviance para llevar a cabo la validación del modelo.

La representación de los residuos de la deviance contra los valores ajustados permite verificar los supuestos del modelo, detectar posibles valores atípicos y determinar si existe un efecto de la variable de grupo para considerar la inclusión de los efectos aleatorios. Dentro de los supuestos del modelo destacan:

- La relación entre el vector de medias y el predictor lineal, expresada mediante la inversa de la función link: $\boldsymbol{\mu} = h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})$.
- La normalidad de los efectos aleatorios: $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$.

Por último, se podría complementar la validación del modelo con la evaluación gráfica de la normalidad de los residuos de la deviance mediante un qq-plot.

2.2.4. Estructura de los modelos del Capítulo 3

A continuación, se explicarán los modelos mixtos que se ajustarán en el [Capítulo 3](#). Aunque los datos se introducirán con más detalle en el [Capítulo 3](#), en esta sección se formulará el modelo de manera genérica. Los datos se agregarán en 36 bloques que se corresponden con unas causas que se explicarán en el [Capítulo 3](#) y se aplicará un modelo mixto a cada uno de ellos. Ahora, se explicará para un bloque siendo de manera análoga para el resto.

Para cada uno de los 36 bloques, la variable respuesta será la matriz \mathbf{Y} de dimensión $m_c \times 1$ con m_c el número de observaciones de cada grupo con $c = 1, \dots, 36$ siendo $m_1 + \dots + m_{36} = n$ donde n es el total de observaciones, la variable explicativa \mathbf{X} , matriz de dimensión $m_c \times 2$, y los grupos vendrán dados por la combinación de dos variables dando lugar a 36 niveles que se explicarán en el [Capítulo 3](#). Como se acaba de comentar, la matriz \mathbf{X} tiene dos columnas: la primera se corresponde con el intercepto y la segunda con la variable numérica a_l con $l = 1, \dots, n_j$ que variará en función de lo que se quiera predecir.

Con todas las variables definidas, se procede a explicar la estructura de los modelos. En concreto, se explica el modelo con intercepto aleatorio que es el que trata de explicar la variable respuesta considerando el efecto de los grupos. La expresión (2.13) de los modelos mixtos se puede reescribir como sigue:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} + \boldsymbol{\varepsilon},$$

equivalentemente,

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & 0 & \dots & 0 \\ \mathbf{X}_2 & 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_J & 0 & 0 & \dots & \mathbf{Z}_J \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_J \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{bmatrix}$$

donde \mathbf{Y}_j es una matriz de dimensión $n_j \times 1$ con $j = 1, \dots, J$ que contiene a todas las observaciones correspondientes al grupo j , \mathbf{Z}_j es una matriz de dimensión $n_j \times 1$ con todo unos y $\boldsymbol{\varepsilon}_j$ es una matriz de dimensión $n_j \times 1$ con todos los errores correspondientes al grupo j . Es decir,

$$\mathbf{Y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_j j} \end{bmatrix} \quad \mathbf{Z}_j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j j} \end{bmatrix}$$

Por otra parte, \mathbf{X}_j es una matriz de dimensión $n_j \times 2$ con la siguiente expresión

$$\mathbf{X}_j = \begin{bmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_{n_j} \end{bmatrix}$$

donde la primera columna tiene todo unos y la segunda incluye la variable numérica introducida anteriormente. Los coeficientes de la regresión serán:

$$\boldsymbol{\beta}' = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}$$

$$\mathbf{b}_j = \begin{bmatrix} b_{0j} \end{bmatrix}$$

donde β_0 es el coeficiente asociado al intercepto y β_1 es el coeficiente asociado a la variable numérica, a_l .

En el [Capítulo 3](#), se van a emplear los modelos explicados en este Capítulo. El último modelo se entenderá mejor cuando se proporcione una descripción detallada de los datos utilizados en el estudio.

2.2.5. Estructura de los modelos del [Apéndice D](#)

Ahora se explicará la estructura de un modelo mixto que se probó pero no dio resultados tan buenos como los anteriores. Los resultados de este se pueden encontrar en el [Apéndice D](#). Los datos se introducirán con más detalle en el [Apéndice D](#), en esta sección se formulará el modelo de manera genérica, empleando tres variables. Aunque se entenderá mejor en el [Apéndice D](#), la variable respuesta será la matriz \mathbf{Y} de dimensión $n \times 1$ con n el número total de observaciones, la explicativa la matriz \mathbf{X} de dimensión $n \times 722$ y los grupos vendrán dados por una variable con 38 niveles que se explicará en el [Apéndice D](#).

En primer lugar, se tiene $\mathbf{t}_{ij} = (t_{ij}^2 \dots t_{ij}^{18})$ que es una matriz de dimensión 1×17 donde cada t_{ij}^l con $l = 2, \dots, 18$ es una variable dicotómica indicadora con la que se sabe a cual de los 18 grupos pertenece la observación y_{ij} . En segundo lugar, se tiene la variable w_{ij} que es una variable dicotómica indicadora de si la observación y_{ij} pertenece o no a un grupo. Por último, está $\mathbf{c}_{ij} = (c_{ij}^2 \dots c_{ij}^{38})$ que es una matriz de dimensión 1×37 donde cada c_{ij}^m con $m = 2, \dots, 38$ es una variable dicotómica indicadora con la que se sabe a cual de los 38 grupos pertenece la observación y_{ij} .

Con estas tres variables definidas, se procede a explicar la estructura de los modelos. En primer lugar, se explica el modelo con intercepto aleatorio que es el que trata de explicar la variable respuesta considerando el efecto de los grupos. Como ya se vio anteriormente, la expresión (2.13) de los modelos mixtos se puede reescribir como sigue:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & 0 & \dots & 0 \\ \mathbf{X}_2 & 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_J & 0 & 0 & \dots & \mathbf{Z}_J \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_J \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_J \end{bmatrix}$$

donde \mathbf{Y}_j es una matriz de dimensión $n_j \times 1$ con $j = 1, \dots, J$ que contiene a todas las observaciones correspondientes al grupo j , \mathbf{Z}_j es una matriz de dimensión $n_j \times 1$ con todo unos y ε_j es una matriz de dimensión $n_j \times 1$ con todos los errores correspondientes al grupo j . Es decir,

$$\mathbf{Y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_j j} \end{bmatrix} \quad \mathbf{Z}_j = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \varepsilon_j = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j j} \end{bmatrix}$$

Por otra parte, \mathbf{X}_j es una matriz de dimensión $n_j \times 722$ con la siguiente expresión

$$\mathbf{X}_j = \begin{bmatrix} 1 & \mathbf{t}_{1j} & w_{1j} & \mathbf{c}_{1j} & \mathbf{t}_{1j}\mathbf{c}_{1j} & \mathbf{w}_{1j}\mathbf{c}_{1j} \\ 1 & \mathbf{t}_{2j} & w_{2j} & \mathbf{c}_{2j} & \mathbf{t}_{2j}\mathbf{c}_{2j} & \mathbf{w}_{2j}\mathbf{c}_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{t}_{n_j j} & w_{n_j j} & \mathbf{c}_{n_j j} & \mathbf{t}_{n_j j}\mathbf{c}_{n_j j} & \mathbf{w}_{n_j j}\mathbf{c}_{n_j j} \end{bmatrix}$$

donde la primera columna tiene todo unos, de la segunda a la decimoctava columnas incluyen los coeficientes de las matrices \mathbf{t}_{ij} , la decimonovena columna es la variable w_{ij} y de la 20 a la 57 contienen

los coeficientes de las matrices \mathbf{c}_{ij} . A partir de la 58 hasta la 687 columna tienen los coeficientes de las matrices $\mathbf{t}_{ij}\mathbf{c}_{ij} = (t_{ij}^2 c_{ij}^2 \dots t_{ij}^2 c_{ij}^{38} \ t_{ij}^3 c_{ij}^2 \dots t_{ij}^3 c_{ij}^{38} \ t_{ij}^{18} c_{ij}^2 \dots t_{ij}^{18} c_{ij}^{38})$ de dimensión 1×629 que representan la interacción entre \mathbf{t}_{ij} y \mathbf{c}_{ij} , cada uno de los coeficientes es una variable indicadora de si la observación y_{ij} pertenece a los dos grupos. Por último, el resto de columnas incluyen los coeficientes de las matrices $\mathbf{w}_{ij}\mathbf{c}_{ij} = (w_{ij}^2 c_{ij}^2 \dots w_{ij}^2 c_{ij}^{38})$ de dimensión 1×37 que representan la interacción entre \mathbf{w}_{ij} y \mathbf{c}_{ij} , cada uno de los coeficientes es una variable indicadora de si la observación y_{ij} pertenece a los dos grupos.

Los coeficientes de la regresión serán:

$$\boldsymbol{\beta}' = \left[\beta_0 \ \beta_1^2 \ \dots \ \beta_1^{18} \ \beta_2 \ \beta_3^2 \ \dots \ \beta_3^{38} \ \beta_4^1 \ \dots \ \beta_4^{629} \ \beta_5^1 \ \dots \ \beta_5^{37} \right]$$

$$\mathbf{b}_j = \left[b_{0j} \right]$$

donde β_0 es el coeficiente asociado al intercepto, β_1^l es el coeficiente asociado a t_{ij}^l , β_2 es el coeficiente asociado a w_{ij} , β_3^m es el coeficiente asociado a c_{ij}^m , β_4^s con $s = 1, \dots, 629$ es el coeficiente asociado a la interacción $t_{ij}^s c_{ij}$, β_5^t con $t = 1, \dots, 37$ es el coeficiente asociado a la interacción $w_{ij} c_{ij}$ y b_{0j} es el coeficiente asociado al intercepto aleatorio.

Para entender mejor la notación con los superíndices, se considera, por ejemplo, el coeficiente β_3^2 . Este coeficiente forma parte del vector $\boldsymbol{\beta}_3$, que está asociado a la matriz \mathbf{c}_{ij} . En este contexto, β_3^2 está asociado a la variable dicotómica c_{ij}^2 , la cual indica si la observación y_{ij} pertenece al grupo 2 de esa variable categórica.

Por último, para el modelo con intercepto y pendiente aleatorios quedaría todo igual excepto las matrices \mathbf{Z}_j y \mathbf{b}_j que en este caso son:

$$\mathbf{Z}_j = \begin{bmatrix} 1 & w_{1j} \\ 1 & w_{2j} \\ \vdots & \vdots \\ 1 & w_{n_j j} \end{bmatrix} \quad \mathbf{b}_j = \begin{bmatrix} b_{0j} \\ b_{1j} \end{bmatrix}$$

donde b_{1j} es el coeficiente asociado a la pendiente aleatoria.

En el [Apéndice D](#), se empleará este modelo y se podrá observar que los resultados no son tan buenos como los que se obtendrán con el modelo del [Capítulo 3](#), ya que con este último se obtienen desviaciones más pequeñas por grupos.

Capítulo 3

Aplicación a los datos

A partir de la Renovación Anual del Padrón se obtuvo la población de las personas residentes en Galicia (población expuesta) y a partir del Registro de Mortalidad de Galicia se obtuvo la mortalidad de la población residente en Galicia durante el período 1999-2022, donde los cuatro últimos años se tomaron como los datos a estudio. En cada uno de estos años se estimó el exceso o defecto de defunciones utilizando como período histórico los diez años previos y se dispuso de la causa básica de defunción codificada con la CIE-10, donde las causas de muerte se encuentran agrupadas en listas de causas (lista reducida) y estas en capítulos (grandes grupos de causas). A partir de esta información de causa de defunción se consideraron los siguientes grupos de causas de interés:

- En primer lugar, se sabe que a parte del capítulo *XVIII*, existen otros códigos que conllevan una mala codificación y se encuentran en capítulos distintos. Por lo tanto, de estos códigos mal definidos, que se pueden ver en el [Capítulo 1](#) de este trabajo, se escogieron las causas con más de veinticinco muertes anuales de media en el período histórico.
- En segundo lugar, entre la clasificación de la lista reducida se seleccionaron las veinte causas con más defunciones. Es decir, se escogieron las veinte causas más frecuentes.
- Y en tercer lugar, se consideraron todos los capítulos de la CIE-10.

De modo que, dentro de los códigos mal definidos finalmente se seleccionaron: *I50.9* (fallo cardíaco agudo), *J96.0* (insuficiencia respiratoria aguda) y *J96.9* (insuficiencia respiratoria, no especificada).

Cuando una de estas causas estaba contenida en una de la lista reducida seleccionada se tuvo en cuenta la causa por una parte y el resto de la lista reducida a la que pertenecía, por otra. Y si esa causa no estaba dentro de la lista reducida seleccionada se tuvo en cuenta la causa por sí misma y el resto de la lista a la que pertenecía se contabilizó como resto del correspondiente capítulo. Del mismo modo se hizo con las causas seleccionadas de la lista reducida. Por ejemplo, una de las causas seleccionadas para estudiar individualmente fue la *I50.9*, incluida dentro de la lista reducida *057*. La lista *057* es una de las veinte más frecuentes, por esta razón esta lista se dividió de la siguiente forma: *057:I50.9* y *057:Resto*. Entonces, el capítulo al que pertenece la lista *057*, que es el capítulo *IX*, se estudia dividido así: *IX:054*, *IX:055*, *IX:056*, *IX:057:I50.9*, *IX:057:Resto*, *IX:058*, *IX:059* y *IX:Resto*. En la [Tabla 3.1](#) se puede ver la definición y el código de la CIE-10 para cada uno de los grupos de causas de defunción seleccionados para el estudio.

Causa de defunción	Definición	Código de la CIE-10
I	Capítulo de enfermedades infecciosas y parasitarias	A00-B99, R75, U04.9, U07.0.1.2, U10.9
II:011 II:012 II:015 II:018 II:028 II:035 II:Resto	Tumor maligno del estómago Tumor maligno del colon Tumor maligno del páncreas Tumor maligno de la tráquea, de los bronquios y del pulmón Tumor maligno de la próstata Tumor maligno de sitios mal definidos, secundarios y de sitios no especificados Resto de causas del capítulo de tumores	C16 C18 C25 C33, C34 C61 C76, C80
III	Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan al mecanismo de la inmunidad	D50-D89
IV:044 IV:Resto	Diabetes mellitus Resto de causas del capítulo enfermedades endocrinas, nutricionales y metabólicas	E10-E14
V:046 V:Resto	Trastornos mentales orgánicos, senil y presenil Resto de causas del capítulo de trastornos mentales y del comportamiento	F00-F09
VI-VIII:051 VI-VIII:052 VI-VIII:Resto	Enfermedad de Alzheimer Otras enfermedades del sistema nervioso y de los órganos de los sentidos Resto de causas del capítulo de Enfermedades del sistema nervioso y de los órganos de los sentidos	G30 Resto G00-H95
IX:054 IX:055 IX:056 IX:057:I50.9 IX:057:Resto IX:058 IX:059 IX:Resto	Enfermedades hipertensivas Infarto agudo de miocardio Otras enfermedades isquémicas del corazón Fallo cardíaco agudo Resto de causas de la lista reducida de insuficiencia cardíaca Otras enfermedades del corazón Enfermedades cerebrovasculares Resto de causas del capítulo de enfermedades del sistema circulatorio	I10-I15 I21 I20, I24-I25 I50.9 I50 (excepto I50.9) I00-I02, I26-I49, I51 I60-I69
X:063 X:064 X:066:J96.0 X:066:J96.9 X:067 X:Resto	Neumonía Enfermedades crónicas de las vías respiratorias inferiores (excepto asma) Insuficiencia respiratoria aguda Insuficiencia respiratoria, no especificada Otras enfermedades del sistema respiratorio Resto de causas del capítulo de enfermedades del sistema respiratorio	J12-J18 J40-J44, J47 J96.0 J96.9 Resto J00-J99
XI:072 XI:Resto	Otras enfermedades del sistema digestivo Resto de causas del capítulo de enfermedades del sistema digestivo	Resto K00-K93
XII	Capítulo de enfermedades de la piel y del tejido subcutáneo	L00-L99
XIII	Capítulo de enfermedades del sistema osteomuscular y del tejido conjuntivo	M00-M99
XIV	Capítulo de enfermedades del sistema genitourinario	N00-N99
XV-XVII	Capítulos: de embarazo, parto y puerperio; de afecciones originadas en el periodo perinatal y de malformaciones congénitas, deformidades y anomalías cromosómicas	O00-O99, P00-P96, Q00-Q99
XVIII	Capítulo de síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte	R00-R74, R76-R99
XX	Capítulo de causas externas de mortalidad	V01-Y89

Tabla 3.1: Definición detallada y código de la CIE-10 para cada grupo de causas de defunción seleccionado.

Es importante destacar que, a pesar de su relevancia, el COVID-19 no está incluido entre los grupos de causas de defunción seleccionados en la [Tabla 3.1](#) debido a no disponer de un histórico. La [Figura 3.1](#) muestra la evolución de la mortalidad observada por COVID-19 en Galicia, con un total de 1617 defunciones en 2020, 1557 en 2021 y 1824 en 2022.

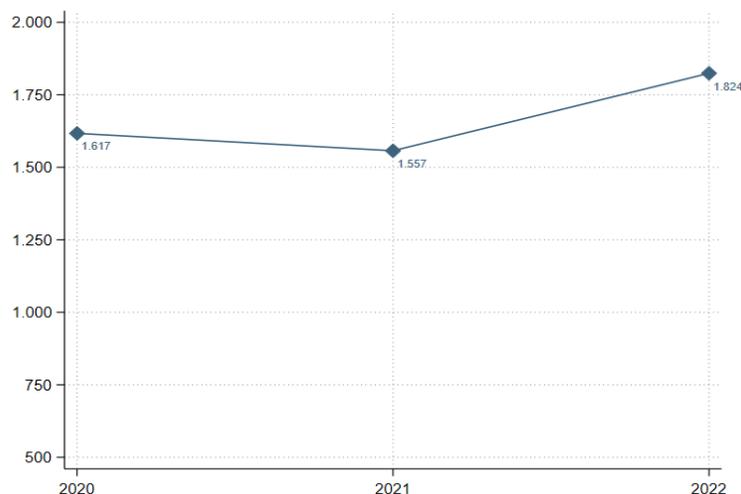


Figura 3.1: Evolución de la mortalidad observada en Galicia por COVID-19 desde el año 2020 al 2022. En el eje X están representados los años y en el eje Y el número de defunciones.

3.1. Modelos lineales generalizados

Se agruparon las defunciones en función del año (considerando el período 1999-2018), el sexo, el grupo de edad (en grupos quinquenales) y la causa de muerte. La variable resultante, número de defunciones, que es una variable de tipo conteo, puede asumirse que sigue una distribución de Poisson o una distribución Binomial Negativa.

La distribución de Poisson es una distribución de probabilidad discreta que expresa la probabilidad de que ocurra un cierto número de sucesos durante un período de tiempo continuo. Como ya se indicó en el [Capítulo 2](#), una de las características más importantes de la distribución de Poisson es que su media y su varianza son iguales, es decir, $\mathbb{E}(Y) = \text{Var}(Y) = \mu$.

Con este enfoque y asumiendo que la variable **defunciones** sigue una distribución de Poisson, se puede aplicar el modelo de regresión de Poisson, introducido en la [Subsección 2.1.1](#), para estimar las defunciones esperadas. El ajuste del modelo se realizó para cada causa seleccionada, se tomó como variable respuesta las **defunciones** y como explicativas el **sexo**, el **grupo de edad** y el **año**, considerando las dos primeras como categóricas y la variable **año** como continua y categórica obteniendo así dos enfoques diferentes, y se tuvo en cuenta la población expuesta.

Sin embargo, un primer análisis descriptivo de las defunciones para cada causa seleccionada y con la estructura de los datos para el modelo hizo pensar que, de forma general, el supuesto de igualdad entre media y varianza no se verificaba ya que en la gran mayoría de casos la varianza fue muy superior a la media, como se observa en la [Tabla 3.2](#). Por tanto, aplicar la regresión de Poisson no sería lo más adecuado debido a la presencia del fenómeno de sobre-dispersión. Por ello, como se ha visto en el [Capítulo 2](#), se optó por emplear el modelo de regresión Binomial Negativa, introducido en la [Subsección 2.1.2](#).

Causa de defunción	Media	Varianza
I	12.77	354.55
II:011	14.98	422.93
II:012	22.84	1130.92
II:015	12.16	266.45
II:018	39.72	3648.78
II:028	14.15	1567.14
II:035	14.42	427.77
II:Resto	107.14	16777.52
III	2.56	67.80
IV:044	15.50	1241.99
IV:Resto	2.85	40.63
V:046	36.58	18589.83
V:Resto	1.82	4.98
VI-VIII:051	16.68	2599.56
VI-VIII:052	15.25	642.13
VI-VIII:Resto	0.28	0.29
IX:054	17.04	2995.24
IX:055	40.62	4533.30
IX:056	33.52	5346.62
IX:057:150.9	33.99	10919.18
IX:057:Resto	12.02	1436.86
IX:058	47.68	13217.17
IX:059	78.69	37811.51
IX:Resto	17.01	1478.91
X:063	17.85	2000.68
X:064	34.76	6136.96
X:066:J96.0	0.82	6.67
X:066:J96.9	2.90	79.12
X:067	40.07	10868.29
X:Resto	3.40	86.82
XI:072	19.30	1750.08
XI:Resto	18.77	583.43
XII	3.33	113.29
XIII	5.35	218.69
XIV	19.78	2746.42
XV-XVII	2.28	43.15
XVIII	18.58	2200.55
XX	36.78	1261.31

Tabla 3.2: Media y varianza de la variable **defunciones** para cada una de las causas seleccionadas, a partir de los datos estructurados para el ajuste del modelo. Galicia, 1999-2018.

Estos modelos asumen que la respuesta sigue una distribución Binomial Negativa que es una generalización de la distribución Geométrica y se define como el número de fracasos antes de obtener el éxito k -ésimo en un experimento de Bernoulli repetido con parámetro p .

El ajuste de este modelo para cada causa seleccionada, se realizó tomando la misma variable respuesta y las mismas variables explicativas que en el caso de la regresión de Poisson. Para comprobar cuál de los dos modelos se ajustó mejor a los datos se llevó a cabo el contraste (2.12), es decir, se tiene en la hipótesis nula el modelo de regresión Poisson y en la alternativa el modelo de regresión Binomial Negativa.

En la [Tabla 3.3](#) se recoge el parámetro de sobre-dispersión para cada causa seleccionada y, como se explicó en el [Capítulo 2](#), cuanto mayor sea el valor de este parámetro, menor será la sobre-dispersión. Estos resultados se obtienen al realizar el ajuste del modelo de la regresión Binomial Negativa en Stata con la función `nbreg`.

En el caso de que la variable `año` se tome como continua, el modelo de regresión Binomial Negativa no converge para las causas *IX:055* y *X:064*. Las causas *V:Resto* y *VI-VIII:Resto* tienen un valor alto del parámetro de sobre-dispersión y un p-valor asociado mayor que los niveles de significación usuales. Además, la causa *II:028*, que aunque no tiene un parámetro de sobre-dispersión tan grande como las anteriores, también presenta un p-valor asociado mayor que los niveles de significación usuales. Por tanto, para estas tres causas no hay evidencias estadísticamente significativas en contra de la hipótesis nula, es decir, no se puede rechazar la idoneidad del modelo de regresión de Poisson. Para las causas restantes, que sí convergen, se tiene que los p-valores son menores que los niveles de significación usuales, por lo tanto hay evidencias estadísticamente significativas en contra de la regresión de Poisson, lo que implicaría que en estas causas existe sobre-dispersión.

Por otro lado, en el caso de que la variable `año` se tome como categórica para la causa *V:046* el modelo de regresión Binomial Negativa no converge. Las causas que tienen un valor alto del parámetro de sobre-dispersión son las mismas que cuando la variable fue considerada como continua y a mayores también están las siguientes causas: *II:028*, *X:066:J96.0*, y *XIII*. El p-valor asociado a todas estas causas es mayor que los niveles de significación usuales. Hay cuatro causas (*IV:Resto*, *VI-VIII:052*, *X:066:J96.9* y *XV-XVII*) que no presentan un parámetro de sobre-dispersión tan grande como las anteriores pero su p-valor es mayor que los niveles de significación usuales. Por último, también están las causas *II:011*, *II:015* y *IX:057:Resto* que tienen asociado un p-valor de 0.02 mayor que 0.01. Por tanto, para todas estas causas no hay evidencias estadísticamente significativas en contra de la hipótesis nula, es decir, no hay nada en contra del modelo de regresión de Poisson.

En conclusión, de forma general, se podría asumir que el modelo de regresión Binomial Negativa es más adecuado que el modelo de Poisson para ajustar el conjunto de los datos.

Causa de defunción	Variable año continua			Variable año categórica		
	Parámetro de sobre-dispersión	Estadístico χ^2	P-valor	Parámetro de sobre-dispersión	Estadístico χ^2	P-valor
I	9.36	337.55	< 0.01	11.46	217.06	< 0.01
II:011	132.49	9.44	< 0.01	170.98	5.84	0.02
II:012	136.31	20.99	< 0.01	251.40	7.09	< 0.01
II:015	103.18	10.54	< 0.01	141.99	5.94	0.02
II:018	14.16	363.91	< 0.01	14.31	346.20	< 0.01
II:028	34.38	0	1	$1.77 \cdot 10^9$	0	1
II:035	41.64	72.29	< 0.01	112.45	14.83	< 0.01
II:Resto	178.67	76.66	< 0.01	217.89	53.99	< 0.01
III	9.27	66.20	< 0.01	21.32	11.66	< 0.01
IV:044	17.28	283.79	< 0.01	36.23	85.38	< 0.01
IV:Resto	37.13	8.08	< 0.01	122.59	0.78	0.38
V:046	30.09	297.73	< 0.01	*	*	*
V:Resto	$1.48 \cdot 10^6$	0	1	$3.22 \cdot 10^6$	0	1
VI-VIII:051	30.17	186.53	< 0.01	60.16	54.73	< 0.01
VI-VIII:052	134.50	8.25	< 0.01	353.64	1.28	0.26
VI-VIII:Resto	$2.85 \cdot 10^{28}$	0	1	$1.99 \cdot 10^{104}$	0	1
IX:054	13.27	362.20	< 0.01	17.36	200.48	< 0.01
IX:055	*	*	*	13.56	950.91	< 0.01
IX:056	14.01	765.43	< 0.01	14.12	744.38	< 0.01
IX:057:I50.9	23.01	301.10	< 0.01	34.23	135.34	< 0.01
IX:057:Resto	23.85	132.39	< 0.01	113.87	6.15	0.02
IX:058	18.95	872.86	< 0.01	19.60	735.39	< 0.01
IX:059	37.56	643.59	< 0.01	39.20	535.71	< 0.01
IX:Resto	16.49	403.87	< 0.01	19.88	299.37	< 0.01
X:063	46.30	87.34	< 0.01	70.99	39.42	< 0.01
X:064	*	*	*	44.10	258.30	< 0.01
X:066:J96.0	4.07	31.33	< 0.01	$2.10 \cdot 10^7$	$7.60 \cdot 10^{-6}$	0.99
X:066:J96.9	4.68	215.27	< 0.01	132.93	0.73	0.39
X:067	25.04	815.80	< 0.01	43.74	273.28	< 0.01
X:Resto	8.69	115.76	< 0.01	33.99	8.18	< 0.01
XI:072	33.98	153.40	< 0.01	37.79	123.06	< 0.01
XI:Resto	22.53	237.62	< 0.01	24.44	206.90	< 0.01
XII	24.79	23.89	< 0.01	44.63	7.20	< 0.01
XIII	73.09	10.31	< 0.01	$2.65 \cdot 10^7$	$3.01 \cdot 10^{-6}$	0.99
XIV	43.40	149.18	< 0.01	53.72	95.73	< 0.01
XV-XVII	14.01	14.13	< 0.01	30.20	1.06	0.30
XVIII	2.78	2858.27	< 0.01	5.63	1797.65	< 0.01
XX	12.57	1334.79	< 0.01	13.48	1170.48	< 0.01

Tabla 3.3: Parámetro de sobre-dispersión, estadístico χ^2 del contraste de sobre-dispersión y su p-valor asociado para cada causa seleccionada y tomando la variable **año** como continua y como categórica. Estos resultados se obtienen al ajustar el modelo de regresión Binomial Negativa en Stata. El símbolo * representa que el modelo no convergió para esa causa.

3.2. Modelos mixtos

En esta sección, se emplearon los modelos mixtos, introducidos en la [Subsección 2.2.4](#), para ajustar los datos, considerando que la variable respuesta `defunciones` sigue una distribución Binomial Negativa. Los detalles sobre cómo se llevaron a cabo los ajustes de los diferentes modelos están recogidos en la [Tabla 3.4](#).

Modelo	Descripción del modelo
Modelo 1	<ul style="list-style-type: none"> ▪ Variable respuesta: <code>defunciones</code>. ▪ Offset: población expuesta.
Modelo 2	<ul style="list-style-type: none"> ▪ Variable respuesta: <code>defunciones</code>. ▪ Offset: población expuesta. ▪ Variable grupo: <code>sexo - grupo de edad</code>.
Modelo 3	<ul style="list-style-type: none"> ▪ Variable respuesta: <code>defunciones</code>. ▪ Offset: población expuesta. ▪ Variable explicativa: <code>año</code>. ▪ Variable grupo: <code>sexo - grupo de edad</code>.
Modelo 4	<ul style="list-style-type: none"> ▪ Variable respuesta: <code>defunciones</code>. ▪ Offset: población expuesta. ▪ Variable explicativa: <code>año</code>. ▪ Variable grupo: <code>sexo - grupo de edad</code>. ▪ Componente <code>ar1</code>.

Tabla 3.4: Descripción de los cuatro modelos mixtos ajustados.

Para cada año a predecir del período 2019-2022, se realizó el ajuste del modelo a partir de los datos observados en la década inmediatamente anterior a cada año. Para el ajuste de los modelos, validaciones y predicciones se usó , en particular, la librería `glmmTMB` que emplea la aproximación de Laplace para realizar el ajuste. Se ajustaron los modelos con  en lugar de Stata debido a que Stata no permite incorporar una componente autorregresiva (`ar1`) a los modelos mixtos, enfoque de interés, al tratar con datos anuales y tener en cuenta su dependencia temporal.

Como la causa *VI-VIII:Resto*, que es la *VI-VIII:050*, tiene pocas defunciones observadas por año se agrupó con la *VI-VIII:052* obteniendo una nueva causa denominada *VI-VIII:050/052*. Pasa lo mismo con la causa *X:066:J96.0* y se agrupó con la *X:066:J96.9* obteniendo la nueva causa *X:066:J96.0.9*. Es decir, en esta sección, en vez de trabajar con 38 causas, finalmente se trabajó con 36.

Al ajustar los modelos para las 36 causas, se observó que, en muchos casos, no se alcanzaba la convergencia. Sin embargo, al cambiar la distribución de Binomial Negativa por la de Poisson, los problemas de convergencia se resolvieron. Por tanto, se ajustaron los datos considerando que la variable `defunciones` seguía una distribución de Poisson.

3.2.1. Año 2019

Empezando por el año 2019, para realizar el ajuste de los modelos se tomó como período histórico de referencia el 2009-2018. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. Es decir, se ajustaron modelos para la variable respuesta `defunciones` considerando solo el intercepto y la población expuesta (Modelo 1 en la [Tabla 3.4](#)) y, por otra parte se ajustaron otros modelos considerando solo el intercepto y la población expuesta pero teniendo en cuenta los efectos aleatorios en función de la combinación entre `sexo` y `grupo de edad` (Modelo 2 en la [Tabla 3.4](#)). Los valores del AIC, que se encuentran en la [Tabla 3.5](#), para el Modelo 1 fueron siempre muy superiores a los del Modelo 2.

Causa de defunción	Modelo 1	Modelo 2	Modelo 3	Modelo 4
I	10308.7	1616.8	1578.7	1562.2
II:011	16300	2346.4	2334.8	1425
II:012	22649.4	1587	1571	1559.3
II:015	10923.6	1394.7	1387.3	1390
II:018	34622.9	1860.6	1862.5	1814.9
II:028	26161.2	685.4	651.7	650.7
II:035	11203.3	1528.8	1479.2	1436.4
II:Resto	74301.4	2457.8	2458.9	2459.2
III	5717.1	1042.7	951.7	924.2
IV:044	19190.4	1389.3	1391.3	1248.2
IV:Resto	4412	1139.6	1115.8	1108
V:046	88478	1395.2	1113.3	1109.2
V:Resto	1754.2	1043.5	1040.5	1043.5
VI-VIII:050/052	17789	1902	1818.8	1822.4
VI-VIII:051	33156.1	931.2	915.9	917
IX:054	32782.2	1312.7	1236.1	1142
IX:055	33649.1	1963.4	1638	1633.3
IX:056	45866	1530.7	1512.4	1497.1
IX:057:I50.9	53929.6	1361.1	1363.1	1353
IX:057:Resto	14540.9	984.9	930	931.3
IX:058	74570.7	1849.1	1848.7	1831.4
IX:059	95275.4	2589.7	1797.1	1788.4
IX:Resto	19829	1261.4	1257.7	1260.4
X:063	27598	1351.1	1311.5	1314.5
X:064	49103.1	1524.2	1333.2	1325.8
X:066:J96.0.9	2802.1	1154.7	694.5	661.1
X:067	60582.5	1999.8	1827.3	1672.7
X:Resto	4854.5	952.6	940.6	918.3
XI:072	26187.2	1509.2	1491.2	1495
XI:Resto	14474.6	1675.2	1615	1609.4
XII	6598.3	733.1	732.5	728.2
XIII	7691.4	1018.7	1006.2	995.9
XIV	36311.2	1301.2	1301	1271.5
XV-XVII	16300	2346.4	2334.8	1020.4
XVIII	13637.8	3658.2	2386.4	1983.1
XX	16300	2346.4	2334.8	2314.2

Tabla 3.5: Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2019. Estos resultados se obtienen con la función AIC de . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.

Ahora, a partir del modelo con efectos aleatorios se ajustó el modelo con intercepto aleatorio tomando como variable explicativa el **año**, considerando que es continua, también se tuvo en cuenta la población expuesta (Modelo 3 en la [Tabla 3.4](#)). Por último, se ajustó el modelo con componente autorregresiva, **ar1**, para incluir la dependencia temporal (Modelo 4 en la [Tabla 3.4](#)). Para la selección de los modelos se escogen los que tengan un menor AIC.

La mayoría de los valores más pequeños del AIC corresponden al Modelo 4, incluso cuando no posee el valor más bajo de AIC, este es muy similar al del mejor modelo que en esos casos sería el Modelo 3 ([Tabla 3.5](#)). Por lo tanto, de forma general para todas las causas se seleccionó el Modelo 4 como el que mejor ajusta los datos. En la [Tabla 3.6](#), están recogidos los coeficientes del Modelo 4 para algunas de las causas.

Causa de defunción	Coefficientes	Estimador	Error típico	Z-value	P-valor
I	Intercepto	96.06	18.72	5.13	< 0.01
	año	-0.05	0.01	-5.67	< 0.01
II:011	Intercepto	0.89	11.12	0.08	0.94
	año	-0.01	0.01	-1.10	0.27
⋮	⋮	⋮	⋮	⋮	⋮
XX	Intercepto	29.22	10.66	2.74	< 0.01
	año	-0.02	0.01	-3.52	< 0.01

Tabla 3.6: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2019.

La validación del modelo se llevó a cabo sobre la causa *I*, obteniéndose resultados muy similares para el resto de las causas. Entonces para la validación del modelo de la causa *I* se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados (gráfico izquierdo de la [Figura 3.2](#)) y el segundo, es el qq-plot de los residuos de la deviance (gráfico derecho de la [Figura 3.2](#)). En el gráfico de la izquierda de la [Figura 3.2](#) se observa que existen patrones no lineales con respecto a los valores ajustados, mientras que en el gráfico de la derecha se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

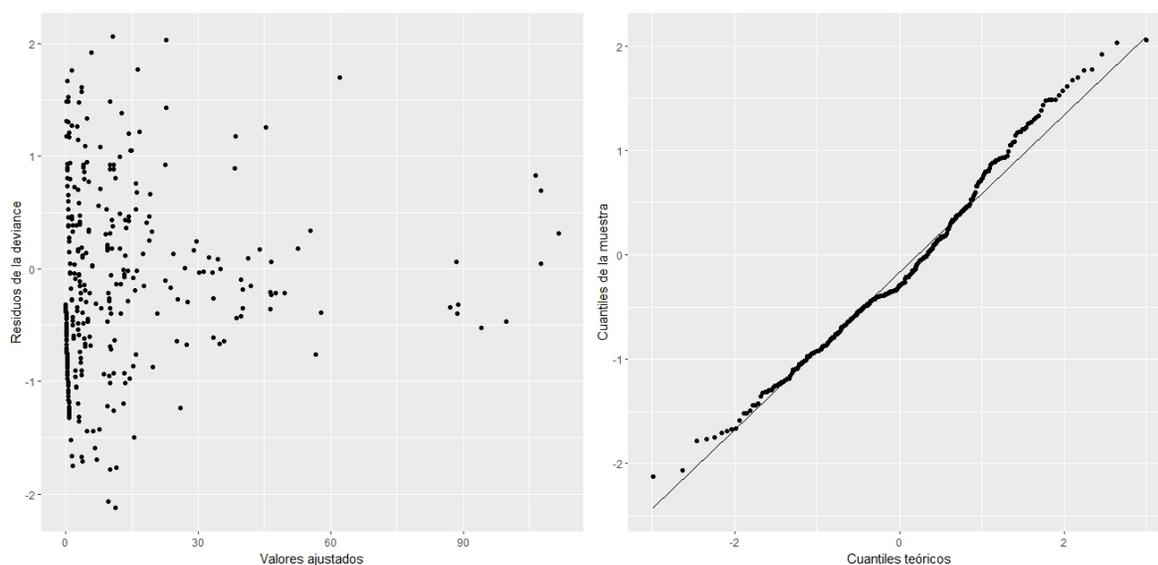


Figura 3.2: Izquierda: Residuos de la deviance del modelo seleccionado para la causa I frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa I . Año 2019.

Tras ajustar el Modelo 4 con cada una de las causas utilizando los datos del período 2009-2018, se obtuvieron las estimaciones anuales del período a estudio y las predicciones para el año 2019. Estas estimaciones y predicciones por causa fueron agregadas para obtener un total anual, estos valores esperados de mortalidad están representados en la [Figura 3.3](#) como puntos rojos donde la escala del eje Y está ajustada al rango 26000-35000 para la mejor visualización de los datos. De modo que, para el año 2019 se identificó un defecto de 1380.6 defunciones, ya que el número de observadas fue de 31184, mientras que la predicción del Modelo 4 fue de 32564.6 defunciones.



Figura 3.3: Evolución de la mortalidad esperada (a partir de los datos del período 2009-2018) y observada en Galicia desde el año 2009 al 2019. En el eje X están representados los años y en el eje Y el número de defunciones.

Para examinar con más detalle el exceso en la predicción del número de defunciones en el año 2019, se han agrupado por el capítulo de la causa de muerte. En la [Tabla 3.7](#), están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura 3.4](#).

Como ya se explicó anteriormente, se obtuvo un defecto de 1380.6 defunciones. En la [Tabla 3.7](#) se observa que los capítulos que más contribuyen a esta sobreestimación son: el capítulo *II* (tumores) con un defecto de 395.4 defunciones, el capítulo *V* (trastornos mentales y de comportamiento) con un defecto de 221.8, el capítulo *IX* (enfermedades del sistema circulatorio) con un defecto de 685 y el capítulo *X* (enfermedades del sistema respiratorio) con un defecto de 476. También cabe destacar que en el capítulo *XVIII* (síntomas/signos/resultados anormales, no clasificados bajo otro concepto) se infraestima, obteniendo así un exceso de 389.3 defunciones en este capítulo lo que supuso un 67.47 % de exceso. El exceso observado en el capítulo *XVIII* se atribuye a que esta situación coincide temporalmente con un proceso de cambio en el personal codificador del Registro de Mortalidad de Galicia. Este proceso implicó la renovación total de la plantilla y la correspondiente formación del nuevo personal.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	412	348.7	63.3	15.36 %
II	8345	8740.4	-395.4	-4.74 %
III	157	206.8	-49.8	-31.72 %
IV	633	715.5	-82.5	-13.03 %
V	2492	2713.8	-221.8	-8.90 %
VI-VIII	1693	1699	-6	-0.35 %
IX	9189	9874	-685	-7.45 %
X	3540	4016	-476	-13.45 %
XI	1559	1556.4	2.6	0.17 %
XII	140	153.5	-13.5	-9.64 %
XIII	169	189.7	-20.7	-12.25 %
XIV	964	919.9	44.1	4.57 %
XV-XVII	64	66.2	-2.2	-3.44 %
XVIII	577	187.7	389.3	67.47 %
XX	1250	1176.9	73.1	5.85 %
Total	31184	32564.6	-1380.6	-4.43 %

Tabla 3.7: Defunciones observadas, esperadas (a partir de los datos del período 2009-2018), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2019.

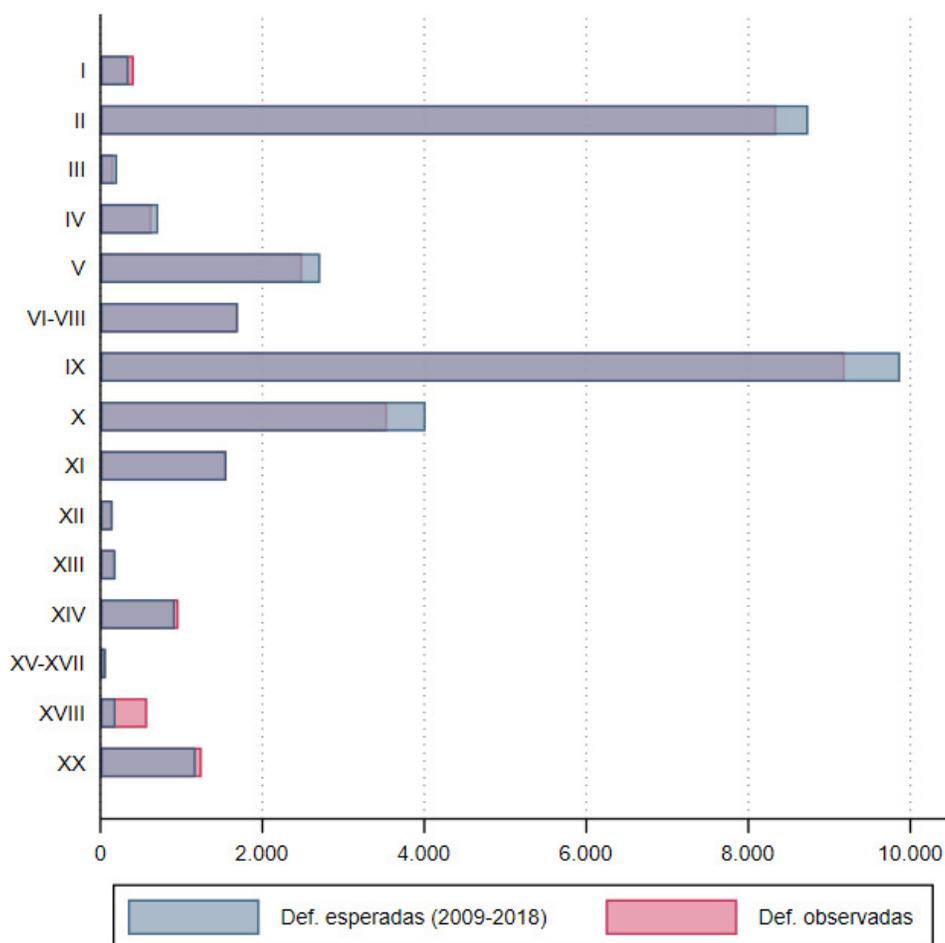


Figura 3.4: Defunciones observadas y esperadas, a partir de los datos del período 2009-2018 en función del capítulo. Galicia, 2019.

Para analizar con mayor detalle lo que ocurre con cada causa de muerte, en la [Tabla 3.8](#) para las 36 causas seleccionadas están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas para el año 2019. Estos valores se presentan gráficamente en la [Figura 3.5](#).

En la [Tabla 3.8](#), se observa que las causas que más contribuyen a la sobreestimación del capítulo *II* (tumores) son la causa *II:011* (tumor maligno del estómago), la *II:018* (tumor maligno de la tráquea, de los bronquios y del pulmón) y la *II:Resto* (resto de causas del capítulo de tumores), con un defecto de 104.1, 99.3 y 108.8 defunciones, respectivamente. En el caso del capítulo *V* (trastornos mentales y de comportamiento), la causa *V:046* (trastornos mentales orgánicos, senil y presenil) es la que más contribuye con un porcentaje de exceso de -8.86% sobre las 2414 defunciones observadas. Para el capítulo *IX* (enfermedades del sistema circulatorio), las causas que tienen un mayor defecto son la *IX:056* (otras enfermedades isquémicas del corazón), la *IX:057:I50.9* (fallo cardíaco agudo), la *IX:058* (otras enfermedades del corazón) y la *IX:Resto* (resto de causas del capítulo de enfermedades del sistema circulatorio), con un defecto de 167.6, 201.5, 144.6 y 96.4 defunciones, respectivamente. Por último, para el capítulo *X* (enfermedades del sistema respiratorio), las causas que más contribuyen al defecto de defunciones son la *X:064* (enfermedades crónicas de las vías respiratorias inferiores, excepto el asma) y la *X:067* (otras enfermedades del sistema respiratorio), con un porcentaje de exceso de -16.49% y -14.78% sobre las 937 y 1830 defunciones observadas, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	412	348.7	63.3	15.36 %
II:011	428	532.1	-104.1	-24.32 %
II:012	836	849.2	-13.2	-1.58 %
II:015	504	571.5	-67.5	-13.39 %
II:018	1500	1599.3	-99.3	-6.62 %
II:028	504	495.6	8.4	1.67 %
II:035	431	442	-11	-2.55 %
II:Resto	4142	4250.8	-108.8	-2.63 %
III	157	206.8	-49.8	-31.72 %
IV:044	471	527.8	-56.8	-12.06 %
IV:Resto	162	187.7	-25.7	-15.86 %
V:046	2414	2628	-214	-8.86 %
V:Resto	78	85.7	-7.7	-9.87 %
VI-VIII:050/052	895	930.7	-35.7	-3.99 %
VI-VIII:051	798	768.2	29.8	3.73 %
IX:054	803	847.1	-44.1	-5.49 %
IX:055	869	955.9	-86.9	-10 %
IX:056	1219	1386.6	-167.6	-13.75 %
IX:057:150.9	1236	1437.5	-201.5	-16.30 %
IX:057:Resto	281	294.5	-13.5	-4.80 %
IX:058	2140	2284.6	-144.6	-6.76 %
IX:059	2142	2072.5	69.5	3.24 %
IX:Resto	499	595.4	-96.4	-19.32 %
X:063	611	652	-41	-6.71 %
X:064	937	1091.5	-154.5	-16.49 %
X:066:J96.0.9	31	6.2	24.8	80 %
X:067	1830	2100.5	-270.5	-14.78 %
X:Resto	131	165.7	-34.7	-26.49 %
XI:072	961	943.8	17.2	1.79 %
XI:Resto	598	612.6	-14.6	-2.44 %
XII	140	153.5	-13.5	-9.64 %
XIII	169	189.7	-20.7	-12.25 %
XIV	964	919.9	44.1	4.57 %
XV-XVII	64	66.2	-2.2	-3.44 %
XVIII	577	187.7	389.3	67.47 %
XX	1250	1176.9	73.1	5.85 %
Total	31184	32564.6	-1380.6	-4.43 %

Tabla 3.8: Defunciones observadas, esperadas (a partir de los datos del período 2009-2018), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2019.

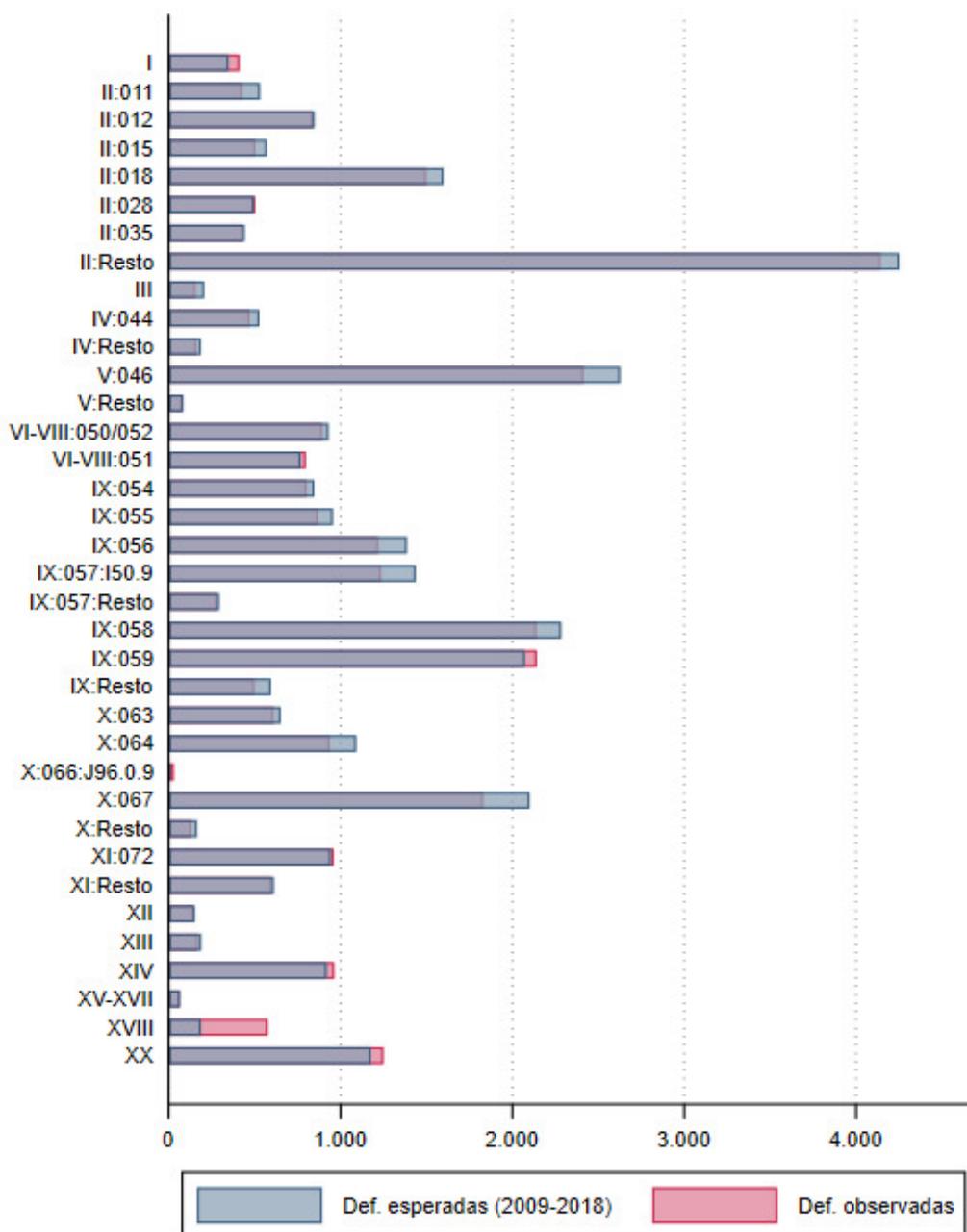


Figura 3.5: Defunciones observadas y esperadas, a partir de los datos del período 2009-2018 en función de las causas de muerte seleccionadas. Galicia, 2019.

3.2.2. Año 2020

A continuación, para el ajuste del modelo del año 2020 se siguió el mismo procedimiento que para el 2019. En este caso, se tomó como período histórico de referencia el 2010-2019. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. Los valores del AIC, que se encuentran en la [Tabla 3.9](#), para el modelo sin efectos aleatorios (Modelo 1 en la [Tabla 3.4](#)) fueron siempre muy superiores a los del modelo con efectos aleatorios (Modelo 2 en la [Tabla 3.4](#)).

Causa de defunción	Modelo 1	Modelo 2	Modelo 3	Modelo 4
I	10287.1	1577.2	1563.2	1550
II:011	16372.4	2290	2291.5	1427.3
II:012	22690.5	1584.3	1538.1	1540.7
II:015	11056.7	1391.2	1392.4	1395.5
II:018	34458.4	1871	1869.9	1820.1
II:028	26229.5	671.1	632.5	635.3
II:035	10703.4	1393.5	1392.6	1395.7
II:Resto	74848.7	2449.8	2445.4	2448
III	6018.6	1025.7	978.5	939.2
IV:044	18292.1	1269.7	1255.5	1196.1
IV:Resto	4650.3	1149.7	1136.6	1128.5
V:046	93345.2	1330	1103.7	1094.3
V:Resto	16372.4	2290	2291.5	1044.4
VI-VIII:050/052	16372.4	2290	2291.5	1858
VI-VIII:051	33470.6	909.5	895.8	895.3
IX:054	33499.5	1285.5	1206.7	1113.1
IX:055	32001.3	1992.8	1610.8	1608.7
IX:056	45422	1569.6	1521.9	1499.4
IX:057:I50.9	54454.3	1379.7	1372.4	1356.9
IX:057:Resto	13977.6	928.2	900.8	903.6
IX:058	75664.9	1878.6	1858	1841.7
IX:059	91477.2	2606.9	1814	1803.2
IX:Resto	19597.4	1304.2	1276.5	1273
X:063	27531.7	1314.6	1269.3	1272.7
X:064	47552.1	1620.2	1334.8	1323.1
X:066:J96.0.9	2005.4	902.1	708.8	645.9
X:067	62942.3	1963.9	1756.6	1635.8
X:Resto	4982.2	935.2	920.2	902.7
XI:072	27343	1501.9	1480.7	1480.4
XI:Resto	14233.3	1659.7	1589.4	1582.7
XII	6725.3	737.7	739.7	736.8
XIII	7517.9	1029.5	1013.8	999.3
XIV	37247.5	1285.2	1280.1	1264.5
XV-XVII	2714.4	1038.9	1037.4	1038.9
XVIII	11960.3	3169.7	2627.6	1983.2
XX	16372.4	2290	2291.5	2280

Tabla 3.9: Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2020. Estos resultados se obtienen con la función AIC de . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.

Ahora, a partir del modelo con efectos aleatorios se ajustó el modelo con intercepto aleatorio tomando como variable explicativa el **año**, considerando que es continua, también se tuvo en cuenta la población expuesta (Modelo 3 en la [Tabla 3.4](#)). Por último, se ajustó el modelo con componente autorregresiva, **ar1**, para incluir la dependencia temporal (Modelo 4 en la [Tabla 3.4](#)). Para la selección de los modelos se escogen los que tengan un menor AIC.

La mayoría de los valores más pequeños del AIC corresponden al Modelo 4. Aunque en algunos casos no tenga el valor más bajo, este es muy similar al del mejor modelo, que suele ser el Modelo 3, con una excepción en la que el mejor modelo es el Modelo 2 ([Tabla 3.9](#)). Por lo tanto, de forma general para todas las causas se seleccionó el Modelo 4 como el que mejor ajusta los datos. En la [Tabla 3.10](#), están recogidos los coeficientes del Modelo 4 para algunas de las causas.

Causa de defunción	Coeficientes	Estimador	Error típico	Z-value	P-valor
I	Intercepto	73.33	20.05	3.66	< 0.01
	año	-0.04	0.01	-4.17	< 0.01
II:011	Intercepto	0.95	51.83	0.02	0.99
	año	-0.01	0.03	-0.24	0.81
⋮	⋮	⋮	⋮	⋮	⋮
XX	Intercepto	5.56	9.51	0.58	0.56
	año	-0.01	0	-1.47	0.14

Tabla 3.10: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2020.

La validación del modelo se llevó a cabo sobre la causa *I*, obteniéndose resultados muy similares para el resto de las causas. Entonces para la validación del modelo de la causa *I* se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados (gráfico izquierdo de la [Figura 3.6](#)) y el segundo, es el qq-plot de los residuos de la deviance (gráfico derecho de la [Figura 3.6](#)). En el gráfico de la izquierda de la [Figura 3.6](#) se observa que existen patrones no lineales con respecto a los valores ajustados, mientras que en el gráfico de la derecha se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

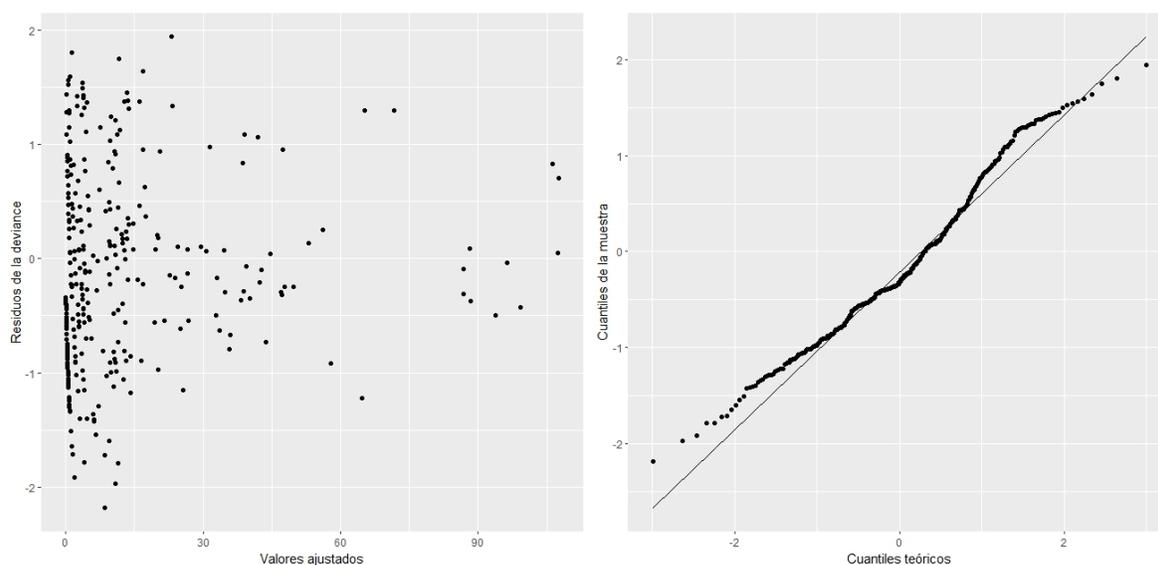


Figura 3.6: Izquierda: Residuos de la deviance del modelo seleccionado para la causa *I* frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa *I*. Año 2020.

Tras ajustar el Modelo 4 con cada una de las causas utilizando los datos del período 2010-2019, se obtuvieron las estimaciones anuales del período a estudio y las predicciones para el año 2020. Estas estimaciones y predicciones por causa fueron agregadas para obtener un total anual, estos valores esperados de mortalidad están representados en la [Figura 3.7](#) como puntos rojos donde la escala del eje Y está ajustada al rango 26000-35000 para la mejor visualización de los datos. De modo que, para el año 2020 se identificó un exceso de 816.9 defunciones, ya que el número de observadas fue de 32823, mientras que la predicción del Modelo 4 fue de 32006.1 defunciones.

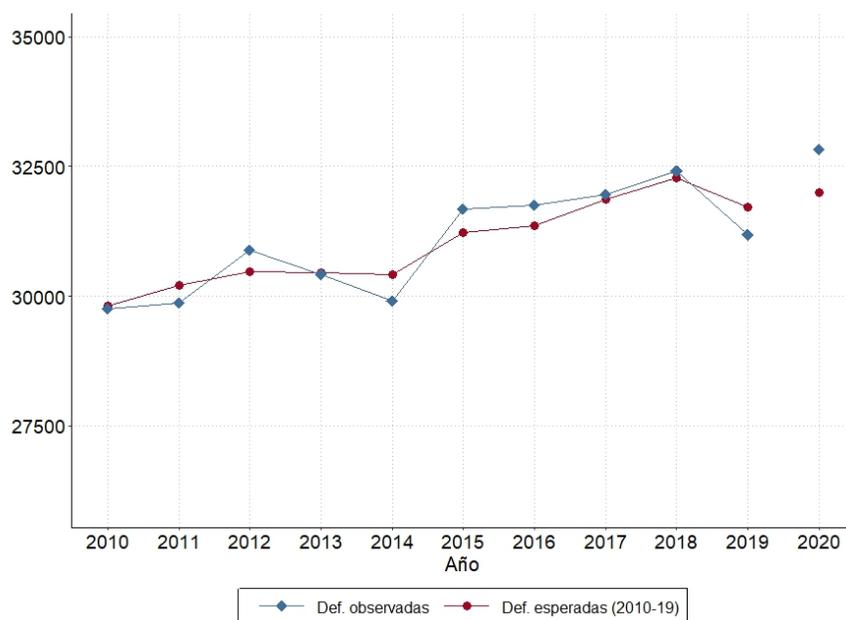


Figura 3.7: Evolución de la mortalidad esperada (a partir de los datos del período 2010-2019) y observada en Galicia desde el año 2010 al 2020. En el eje X están representados los años y en el eje Y el número de defunciones.

Para examinar con más detalle el exceso en la predicción del número de defunciones en el año 2020, se han agrupado por el capítulo de la causa de muerte. En la [Tabla 3.11](#), están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura 3.8](#).

Como se mencionó anteriormente, se identificó un exceso de 816.9 defunciones para el total anual. Al observar la [Tabla 3.11](#), destaca con diferencia el capítulo *I* (enfermedades infecciosas y parasitarias) como el que más contribuye a la infraestimación del año 2020, con un exceso de 1623.8 defunciones. Esta situación se debe a que el COVID-19 no existía en el período 2010-2019. También se infraestima en el capítulo *XVIII* (síntomas/signos/resultados anormales, no clasificados bajo otro concepto) obteniendo así un porcentaje de exceso de 38.70 % sobre las 391 defunciones observadas. Esto se debe a que durante el 2020, la calidad de las causas básicas de defunción pudo haberse visto afectada debido a la adaptación del personal recientemente incorporado en el Registro de Mortalidad de Galicia. Por otro lado, tanto en el capítulo *V* (trastornos mentales y de comportamiento) como en el *X* (enfermedades del sistema respiratorio) se observa una sobreestimación mayor con respecto a los demás capítulos, con un defecto de 415.8 y 549.8 defunciones, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	1988	364.2	1623.8	81.68 %
II	8483	8603.3	-120.3	-1.42 %
III	168	189.7	-21.7	-12.92 %
IV	762	732.1	29.9	3.92 %
V	2439	2854.8	-415.8	-17.05 %
VI-VIII	1721	1616.8	104.2	6.05 %
IX	9505	9605.7	-100.7	-1.06 %
X	3139	3688.8	-549.8	-17.52 %
XI	1541	1545.4	-4.4	-0.29 %
XII	149	144.6	4.4	2.95 %
XIII	160	171.6	-11.6	-7.25 %
XIV	1016	940.6	75.4	7.42 %
XV-XVII	72	74.3	-2.3	-3.19 %
XVIII	391	239.7	151.3	38.70 %
XX	1289	1234.5	54.5	4.23 %
Total	32823	32006.1	816.9	2.49 %

Tabla 3.11: Defunciones observadas, esperadas (a partir de los datos del período 2010-2019), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2020.

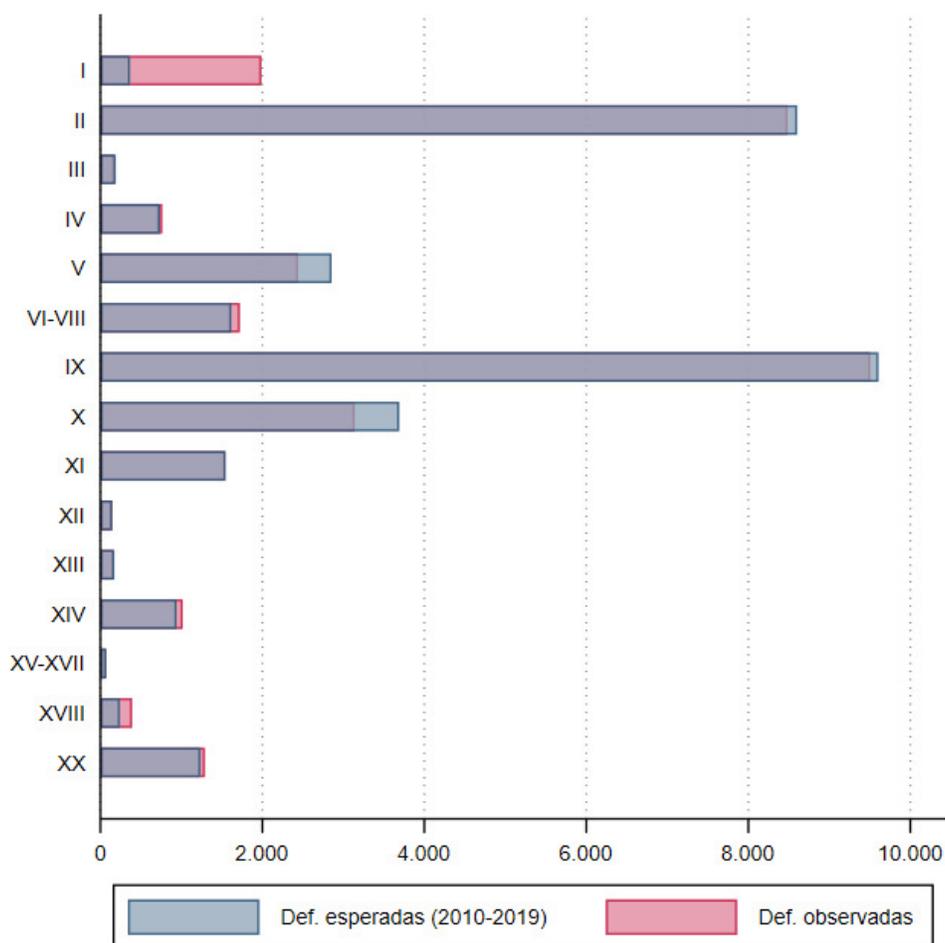


Figura 3.8: Defunciones observadas y esperadas, a partir de los datos del período 2010-2019 en función del capítulo. Galicia, 2020.

Para analizar con mayor profundidad lo que ocurre con cada causa de muerte, en la [Tabla 3.12](#) para las 36 causas seleccionadas están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas para el año 2020. Estos valores se presentan gráficamente en la [Figura 3.9](#).

En la [Tabla 3.12](#), se observa que la causa que más contribuye a la infraestimación del capítulo *V* (trastornos mentales y de comportamiento) es la causa *V:046* (trastornos mentales orgánicos, senil y presenil) con un porcentaje de exceso de -18.21% sobre las 2351 defunciones observadas. Por otro lado, para el capítulo *X* (enfermedades del sistema respiratorio), las causas que más contribuyen al defecto de defunciones son la *X:063* (neumonía), la *X:064* (enfermedades crónicas de las vías respiratorias inferiores, excepto el asma) y la *X:067* (otras enfermedades del sistema respiratorio), con un defecto de 143.4, 144.8 y 253.6 defunciones, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	1988	364.2	1623.8	81.68 %
II:011	459	508.9	-49.9	-10.87 %
II:012	834	840.6	-6.6	-0.79 %
II:015	562	544.9	17.1	3.04 %
II:018	1540	1551.1	-11.1	-0.72 %
II:028	478	478.9	-0.9	-0.19 %
II:035	434	431.4	2.6	0.60 %
II:Resto	4176	4247.4	-71.4	-1.71 %
III	168	189.7	-21.7	-12.92 %
IV:044	569	554.4	14.6	2.57 %
IV:Resto	193	177.7	15.3	7.93 %
V:046	2351	2779	-428	-18.21 %
V:Resto	88	75.8	12.2	13.86 %
VI-VIII:050/052	926	839.5	86.5	9.34 %
VI-VIII:051	795	777.2	17.8	2.24 %
IX:054	1097	899.2	197.8	18.03 %
IX:055	846	878.8	-32.8	-3.88 %
IX:056	1256	1311.1	-55.1	-4.39 %
IX:057:150.9	1325	1398.1	-73.1	-5.52 %
IX:057:Resto	347	279.4	67.6	19.48 %
IX:058	2075	2260	-185	-8.92 %
IX:059	2055	2038.8	16.2	0.79 %
IX:Resto	504	540.4	-36.4	-7.22 %
X:063	496	639.4	-143.4	-28.91 %
X:064	842	986.8	-144.8	-17.20 %
X:066:J96.0.9	55	12.5	42.5	77.27 %
X:067	1659	1912.6	-253.6	-15.29 %
X:Resto	87	137.5	-50.5	-58.05 %
XI:072	929	966.5	-37.5	-4.04 %
XI:Resto	612	578.9	33.1	5.41 %
XII	149	144.6	4.4	2.95 %
XIII	160	171.6	-11.6	-7.25 %
XIV	1016	940.6	75.4	7.42 %
XV-XVII	72	74.3	-2.3	-3.19 %
XVIII	391	239.7	151.3	38.70 %
XX	1289	1234.5	54.5	4.23 %
Total	32823	32006.1	816.9	2.49 %

Tabla 3.12: Defunciones observadas, esperadas (a partir de los datos del período 2010-2019), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2020.

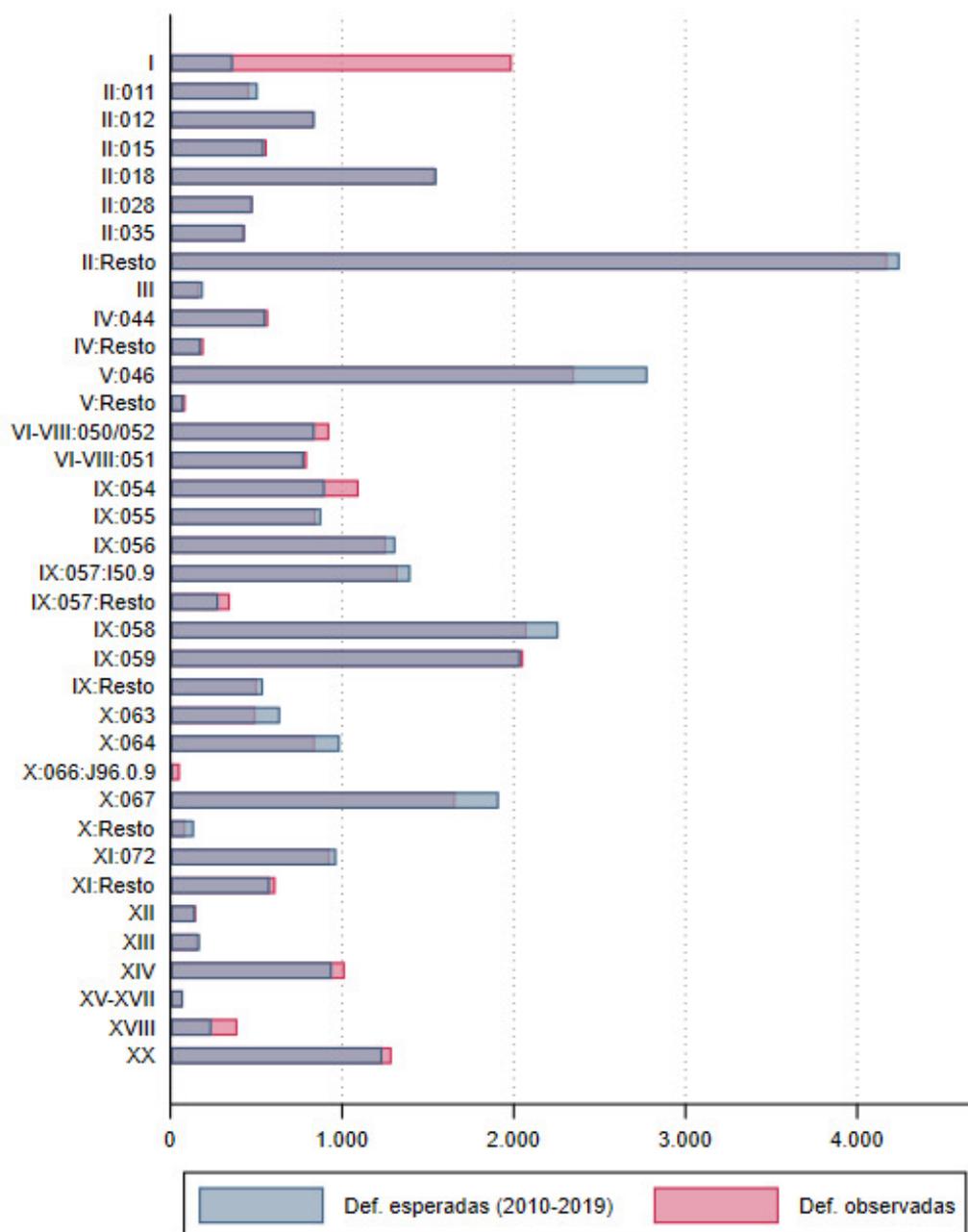


Figura 3.9: Defunciones observadas y esperadas, a partir de los datos del período 2010-2019 en función de la de las causas de muerte seleccionadas. Galicia, 2020.

3.2.3. Año 2021

Ahora, para el ajuste del modelo del año 2021 se siguieron los mismos pasos que en los dos años anteriores. En este caso, se tomó como período histórico de referencia el 2011-2020. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. Los valores del AIC, que se encuentran en la [Tabla 3.13](#), para el modelo sin efectos aleatorios (Modelo 1 en la [Tabla 3.4](#)) fueron siempre muy superiores a los del modelo con efectos aleatorios (Modelo 2 en la [Tabla 3.4](#)).

Causa de defunción	Modelo 1	Modelo 2	Modelo 3	Modelo 4
I	18484.6	4229.2	3355.3	1778.5
II:011	16653.3	2272.9	2274.5	1421
II:012	22500.1	1583.5	1528.2	1532.2
II:015	11220.1	1384.9	1386.5	1388.5
II:018	34240.9	1906.5	1898.2	1829.3
II:028	25959.2	657.6	620.6	624.5
II:035	10659.3	1395.7	1395.9	1398.2
II:Resto	75273.2	2464.8	2450.3	2446.5
III	6319.9	1030.9	1010	969.4
IV:044	18846	1263.3	1255.8	1199
IV:Resto	4996.9	1153.6	1129.9	1132.2
V:046	97552.4	1238.7	1130	1104.4
V:Resto	1811.4	1040.7	1042	1044.7
VI-VIII:050/052	16653.3	2272.9	2274.5	1861
VI-VIII:051	33750.9	880.4	870.5	873.1
IX:054	35689.7	1311.9	1226.5	1144.2
IX:055	30089.3	2033.9	1621.4	1617.8
IX:056	44703.5	1614.8	1550.9	1510.5
IX:057:I50.9	55149.1	1369.1	1349.9	1331.2
IX:057:Resto	14133.8	943.4	916.2	919.8
IX:058	75586.1	1923.3	1870.8	1856.3
IX:059	87702.6	2554.3	1788.6	1771.6
IX:Resto	19243.3	1342	1290.7	1282.8
X:063	26597.2	1358.2	1268.8	1259.2
X:064	45232.4	1729.4	1349.9	1322.2
X:066:J96.0.9	1523.3	710.9	699.1	623.3
X:067	64074.4	1958.5	1826	1623.8
X:Resto	4859.9	955.5	952.5	911.9
XI:072	16653.3	2272.9	2274.5	1461.3
XI:Resto	14205.7	1653.4	1567.5	1559.5
XII	6896.2	731.5	731.9	721.7
XIII	7152.4	1016	1004.8	994.6
XIV	38227	1248.6	1242	1240.2
XV-XVII	16653.3	2272.9	2274.5	1026.1
XVIII	11300.9	2814.2	2645.9	1906.7
XX	16653.3	2272.9	2274.5	2268.6

Tabla 3.13: Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2021. Estos resultados se obtienen con la función AIC de . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.

A partir del modelo con efectos aleatorios se ajustó el modelo con intercepto aleatorio tomando como variable explicativa el **año**, considerando que es continua, también se tuvo en cuenta la población expuesta (Modelo 3 en la [Tabla 3.4](#)). Por último, se ajustó el modelo con componente autorregresiva, **ar1**, para incluir la dependencia temporal (Modelo 4 en la [Tabla 3.4](#)). Para la selección de los modelos se escogen los que tengan un menor AIC.

La mayoría de los valores más pequeños del AIC corresponden al Modelo 4, aunque en algunos casos no tenga el valor más bajo. Sin embargo, estos valores son muy similares a los del mejor modelo en esos casos, que sería el Modelo 3 o el Modelo 2 ([Tabla 3.13](#)). Por lo tanto, de forma general para todas las causas se seleccionó el Modelo 4 como el que mejor ajusta los datos. En la [Tabla 3.14](#), están recogidos los coeficientes del Modelo 4 para algunas de las causas.

Causa de defunción	Coefficientes	Estimador	Error típico	Z-value	P-valor
I	Intercepto	-53.07	35.13	-1.51	0.13
	año	0.02	0.02	1.22	0.22
II:011	Intercepto	-1.11	91.91	-0.01	0.99
	año	-0.01	0.05	-0.11	0.91
⋮	⋮	⋮	⋮	⋮	⋮
XX	Intercepto	-0.01	0.05	-0.11	0.91
	año	0	0	-0.13	0.90

Tabla 3.14: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2021.

La validación del modelo se llevó a cabo sobre la causa *I*, obteniéndose resultados muy similares para el resto de las causas. Entonces para la validación del modelo de la causa *I* se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados (gráfico izquierdo de la [Figura 3.10](#)) y el segundo, es el qq-plot de los residuos de la deviance (gráfico derecho de la [Figura 3.10](#)). En el gráfico de la izquierda de la [Figura 3.10](#) se observa que existen patrones no lineales con respecto a los valores ajustados, mientras que en el gráfico de la derecha se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

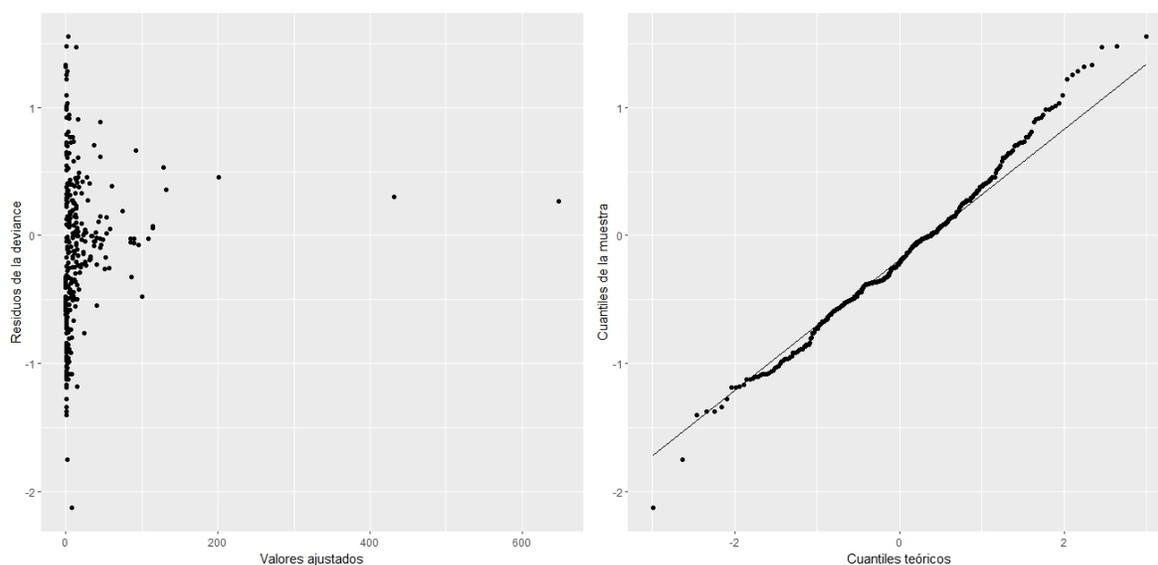


Figura 3.10: Izquierda: Residuos de la deviance del modelo seleccionado para la causa *I* frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa *I*. Año 2021.

Tras ajustar el Modelo 4 con cada una de las causas utilizando los datos del período 2011-2020, se obtuvieron las estimaciones anuales del período a estudio y las predicciones para el año 2021. Estas estimaciones y predicciones por causa fueron agregadas para obtener un total anual, estos valores esperados de mortalidad están representados en la [Figura 3.11](#) como puntos rojos donde la escala del eje Y está ajustada al rango 26000-35000 para la mejor visualización de los datos. De modo que, para el año 2021 se identificó un exceso de 412.8 defunciones, ya que el número de observadas fue de 32853, mientras que la predicción del Modelo 4 fue de 32440.2 defunciones.

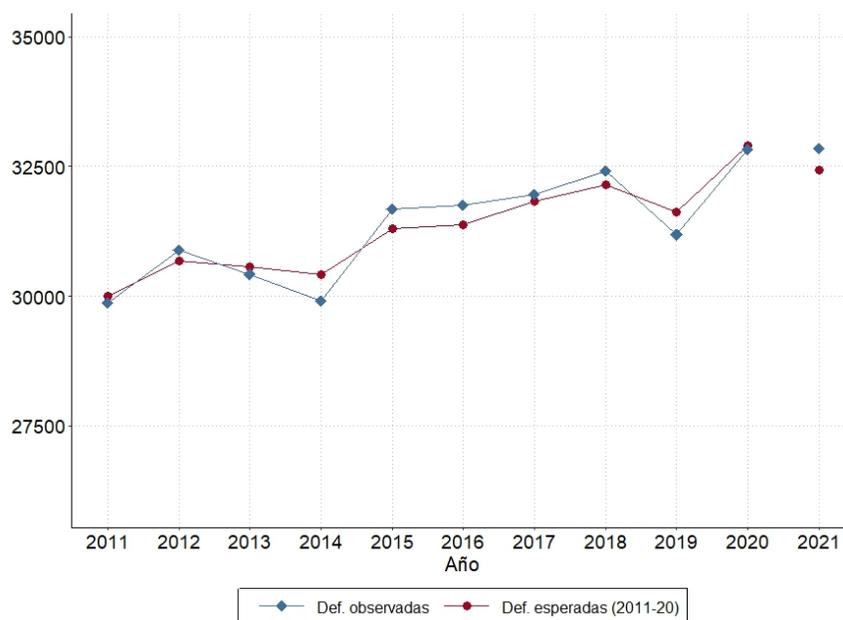


Figura 3.11: Evolución de la mortalidad esperada (a partir de los datos del período 2011-2020) y observada en Galicia desde el año 2011 al 2021. En el eje *X* están representados los años y en el eje *Y* el número de defunciones.

Para examinar con más detalle el exceso en la predicción del número de defunciones en el año 2021, se han agrupado por el capítulo de la causa de muerte. En la [Tabla 3.15](#), están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura 3.12](#).

Como se comentó anteriormente, se identificó un exceso de 412.8 defunciones para el total anual. En la [Tabla 3.15](#), destaca con diferencia el capítulo *I* (enfermedades infecciosas y parasitarias) como el que más contribuye a la infraestimación del año 2020, con un exceso de 967.6 defunciones. En este caso, aunque el COVID-19 ya existía, solo había registros del año 2020. Además, el capítulo *XVIII* (síntomas/signos/resultados anormales, no clasificados bajo otro concepto) también infraestima con un exceso de 340.8 defunciones. Este exceso observado en el capítulo *XVIII* puede atribuirse a la continuidad en el proceso de adaptación del personal codificador. Por otro lado, tanto en el capítulo *V* (trastornos mentales y de comportamiento) como en el *X* (enfermedades del sistema respiratorio) se observa una sobreestimación mayor con respecto a los demás capítulos, con un defecto de 550.2 y 781.4 defunciones, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	1954	986.4	967.6	49.52 %
II	8464	8543.7	-79.7	-0.94 %
III	168	168.4	-0.4	-0.24 %
IV	900	781.6	118.4	13.16 %
V	2227	2777.2	-550.2	-24.71 %
VI-VIII	1573	1659.3	-86.3	-5.49 %
IX	9581	9520.3	60.7	0.63 %
X	2885	3666.4	-781.4	-27.08 %
XI	1589	1571.3	17.7	1.11 %
XII	148	134	14	9.46 %
XIII	220	169.5	50.5	22.95 %
XIV	1128	974.1	153.9	13.64 %
XV-XVII	88	63.7	24.3	27.61 %
XVIII	488	147.2	340.8	69.84 %
XX	1440	1277.3	162.7	11.30 %
Total	32853	32440.2	412.8	1.26 %

Tabla 3.15: Defunciones observadas, esperadas (a partir de los datos del período 2011-2020), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2021.

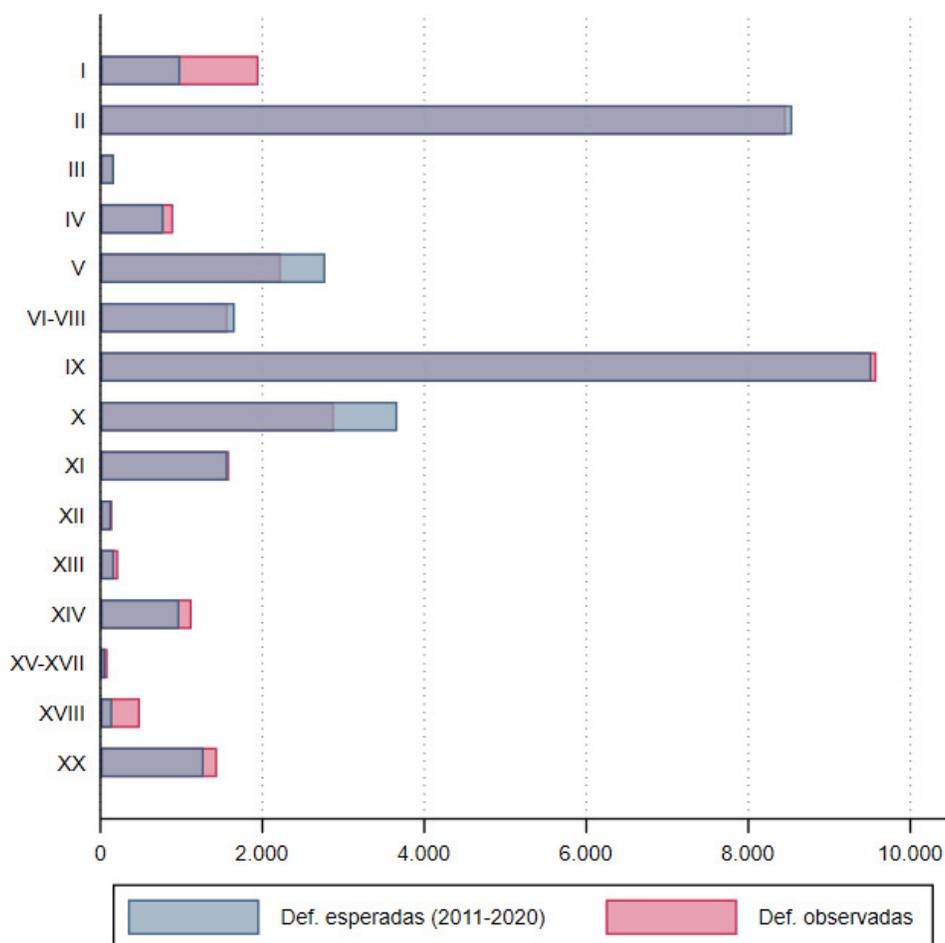


Figura 3.12: Defunciones observadas y esperadas, a partir de los datos del período 2011-2020 en función del capítulo. Galicia, 2021.

Para analizar con mayor profundidad lo que ocurre con cada causa de muerte, en la [Tabla 3.16](#) para las 36 causas seleccionadas están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas para el año 2021. Estos valores se presentan gráficamente en la [Figura 3.13](#).

En la [Tabla 3.16](#), se observa que la causa que más contribuye a la infraestimación del capítulo *V* (trastornos mentales y de comportamiento) es la causa *V:046* (trastornos mentales orgánicos, senil y presenil) con un porcentaje de exceso de -25.01% sobre las 2154 defunciones observadas. Por otro lado, para el capítulo *X* (enfermedades del sistema respiratorio), las causas que más contribuyen al defecto de defunciones son la *X:063* (neumonía), la *X:064* (enfermedades crónicas de las vías respiratorias inferiores, excepto el asma) y la *X:067* (otras enfermedades del sistema respiratorio), con un defecto de 141.1, 128.1 y 553.5 defunciones, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	1954	986.4	967.6	49.52 %
II:011	379	503.2	-124.2	-32.77 %
II:012	822	827.4	-5.4	-0.66 %
II:015	521	557.7	-36.7	-7.04 %
II:018	1586	1533.9	52.1	3.28 %
II:028	496	479.4	16.6	3.35 %
II:035	444	439.8	4.2	0.95 %
II:Resto	4216	4202.2	13.8	0.33 %
III	168	168.4	-0.4	-0.24 %
IV:044	618	592.2	25.8	4.17 %
IV:Resto	282	189.4	92.6	32.84 %
V:046	2154	2692.8	-538.8	-25.01 %
V:Resto	73	84.4	-11.4	-15.62 %
VI-VIII:050/052	906	870.3	35.7	3.94 %
VI-VIII:051	667	789	-122	-18.29 %
IX:054	1120	1131.6	-11.6	-1.04 %
IX:055	851	815.2	35.8	4.21 %
IX:056	1188	1318.1	-130.1	-10.95 %
IX:057:150.9	1441	1300.3	140.7	9.76 %
IX:057:Resto	360	293.5	66.5	18.47 %
IX:058	2075	2140.1	-65.1	-3.14 %
IX:059	2056	2000.8	55.2	2.68 %
IX:Resto	490	520.8	-30.8	-6.29 %
X:063	443	584.1	-141.1	-31.85 %
X:064	749	877.1	-128.1	-17.10 %
X:066:J96.0.9	94	36.7	57.3	60.96 %
X:067	1505	2058.5	-553.5	-36.78 %
X:Resto	94	109.9	-15.9	-16.91 %
XI:072	974	990.7	-16.7	-1.71 %
XI:Resto	615	580.6	34.4	5.59 %
XII	148	134	14	9.46 %
XIII	220	169.5	50.5	22.95 %
XIV	1128	974.1	153.9	13.64 %
XV-XVII	88	63.7	24.3	27.61 %
XVIII	488	147.2	340.8	69.84 %
XX	1440	1277.3	162.7	11.30 %
Total	32853	32440.2	412.8	1.26 %

Tabla 3.16: Defunciones observadas, esperadas (a partir de los datos del período 2011-2020), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2021.

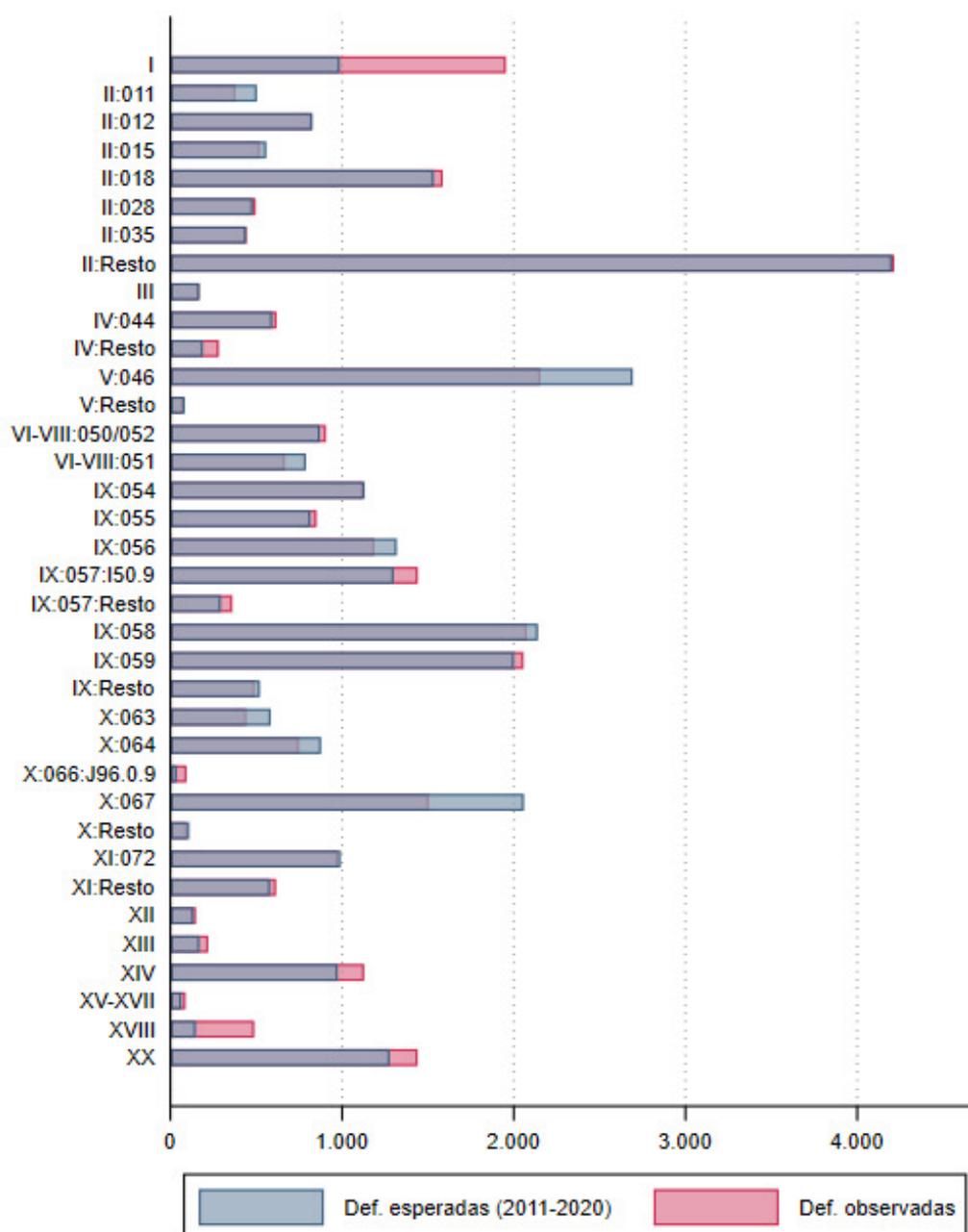


Figura 3.13: Defunciones observadas y esperadas, a partir de los datos del período 2011-2020 en función de las causas de muerte seleccionadas. Galicia, 2021.

3.2.4. Año 2022

Por último, para el ajuste del modelo del año 2022 se siguió el mismo procedimiento que en los años anteriores. En este caso, se tomó como período histórico de referencia el 2011-2021. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. Los valores del AIC, que se encuentran en la [Tabla 3.17](#), para el modelo sin efectos aleatorios (Modelo 1 en la [Tabla 3.4](#)) fueron siempre muy superiores a los del modelo con efectos aleatorios (Modelo 2 en la [Tabla 3.4](#)).

Causa de defunción	Modelo 1	Modelo 2	Modelo 3	Modelo 4
I	25119.9	5735	3418.8	1807
II:011	11571.8	1442.9	1376.7	1380
II:012	22264.7	1578.6	1519.7	1523.4
II:015	11200.5	1397.8	1399.7	1401.4
II:018	33885	1907.9	1893.1	1814.5
II:028	25511.7	644.8	620.2	623.7
II:035	10729.2	1378.8	1380.7	1383.2
II:Resto	75923.9	2479.7	2452.5	2448.8
III	6632.9	996.3	996.4	964.4
IV:044	19490.5	1245.4	1238.7	1184.6
IV:Resto	5607	1229.4	1175.1	1160.4
V:046	99870.6	1220.9	1201.5	1123.4
V:Resto	1823	1033	1035	1037.9
VI-VIII:050/052	20406.4	1886.1	1825.9	1828.8
VI-VIII:051	33081.9	919.3	896	890.2
IX:054	37967.9	1293.1	1229	1162.3
IX:055	28283.7	1997.2	1586.8	1584.8
IX:056	43064	1684	1602.4	1547.6
IX:057:I50.9	17317	2299.1	2292.5	1343.9
IX:057:Resto	14207.1	948.6	934.6	936.6
IX:058	75657.9	1955.3	1848.8	1840.7
IX:059	84369.1	2503.7	1772.5	1752.6
IX:Resto	18946.1	1373.6	1287.4	1281.8
X:063	25791.7	1421	1248.6	1226.7
X:064	42678.6	1827.3	1352	1314.1
X:066:J96.0.9	1492.1	682.6	609.5	591.1
X:067	65236.1	1962.7	1956.3	1634.7
X:Resto	4756.4	966	967	915.8
XI:072	17317	2299.1	2292.5	1469.9
XI:Resto	13953.5	1643	1562.2	1551.3
XII	17317	2299.1	2292.5	716.9
XIII	7057.1	971.7	973.7	971.7
XIV	39533.5	1248.3	1246.5	1240.1
XV-XVII	17317	2299.1	2292.5	1055.9
XVIII	11230.3	2467.7	2468.7	1821
XX	17317	2299.1	2292.5	2279.8

Tabla 3.17: Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2022. Estos resultados se obtienen con la función AIC de . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.

A partir del modelo con efectos aleatorios se ajustó el modelo con intercepto aleatorio tomando como variable explicativa el **año**, considerando que es continua, también se tuvo en cuenta la población expuesta (Modelo 3 en la [Tabla 3.4](#)). Por último, se ajustó el modelo con componente autorregresiva, **ar1**, para incluir la dependencia temporal (Modelo 4 en la [Tabla 3.4](#)). Para la selección de los modelos se escogen los que tengan un menor AIC.

La mayoría de los valores más pequeños del AIC corresponden al Modelo 4, aunque en algunos casos no tenga el valor más bajo. Sin embargo, estos valores son muy similares a los del mejor modelo en esos casos, que sería el Modelo 3 o el Modelo 2 ([Tabla 3.17](#)). Por lo tanto, de forma general para todas las causas se seleccionó el Modelo 4 como el que mejor ajusta los datos. En la [Tabla 3.18](#), están recogidos los coeficientes del Modelo 4 para algunas de las causas.

Causa de defunción	Coefficientes	Estimador	Error típico	Z-value	P-valor
I	Intercepto	-122.25	39.41	-3.10	< 0.01
	año	0.06	0.02	2.84	< 0.01
II:011	Intercepto	74	11	7	< 0.01
	año	0	0	-8	< 0.01
⋮	⋮	⋮	⋮	⋮	⋮
XX	Intercepto	-22.21	9.81	-2.26	0.02
	año	0.01	0	1.40	0.16

Tabla 3.18: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2022.

La validación del modelo se llevó a cabo sobre la causa *I*, obteniéndose resultados muy similares para el resto de las causas. Entonces para la validación del modelo de la causa *I* se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados (gráfico izquierdo de la [Figura 3.14](#)) y el segundo, es el qq-plot de los residuos de la deviance (gráfico derecho de la [Figura 3.14](#)). En el gráfico de la izquierda de la [Figura 3.14](#) se observa que existen patrones no lineales con respecto a los valores ajustados, mientras que en el gráfico de la derecha se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

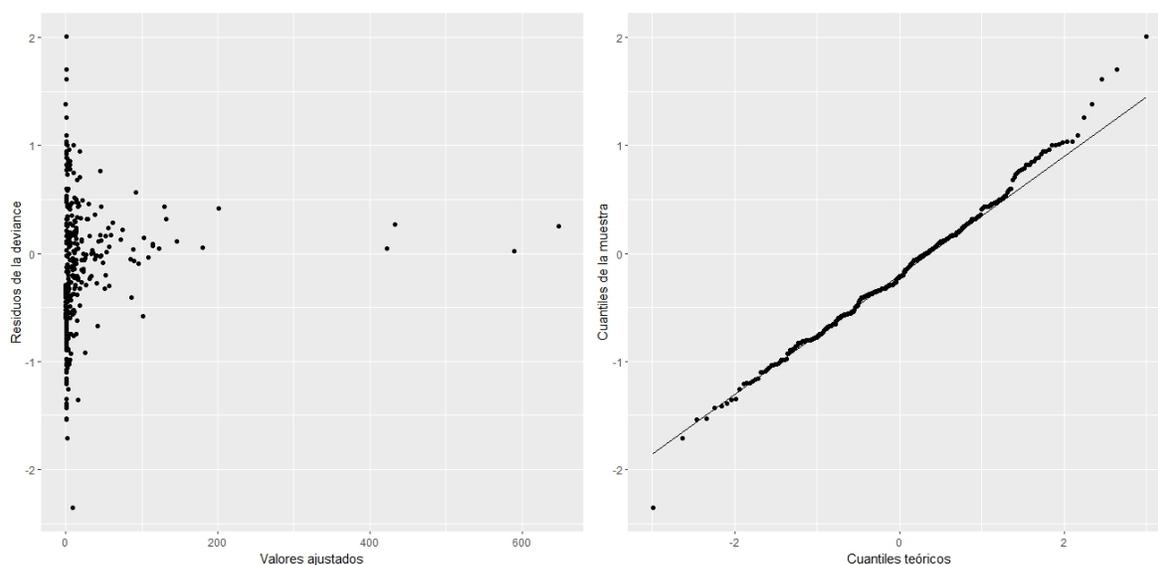


Figura 3.14: Izquierda: Residuos de la deviance del modelo seleccionado para la causa *I* frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa *I*. Año 2022.

Tras ajustar el Modelo 4 con cada una de las causas utilizando los datos del período 2012-2021, se obtuvieron las estimaciones anuales del período a estudio y las predicciones para el año 2022. Estas estimaciones y predicciones por causa fueron agregadas para obtener un total anual, estos valores esperados de mortalidad están representados en la [Figura 3.15](#) como puntos rojos donde la escala del eje Y está ajustada al rango 26000-35000 para la mejor visualización de los datos. De modo que, para el año 2022 se identificó un exceso de 2111.7 defunciones, ya que el número de observadas fue de 34487, mientras que la predicción del Modelo 4 fue de 32375.3 defunciones.

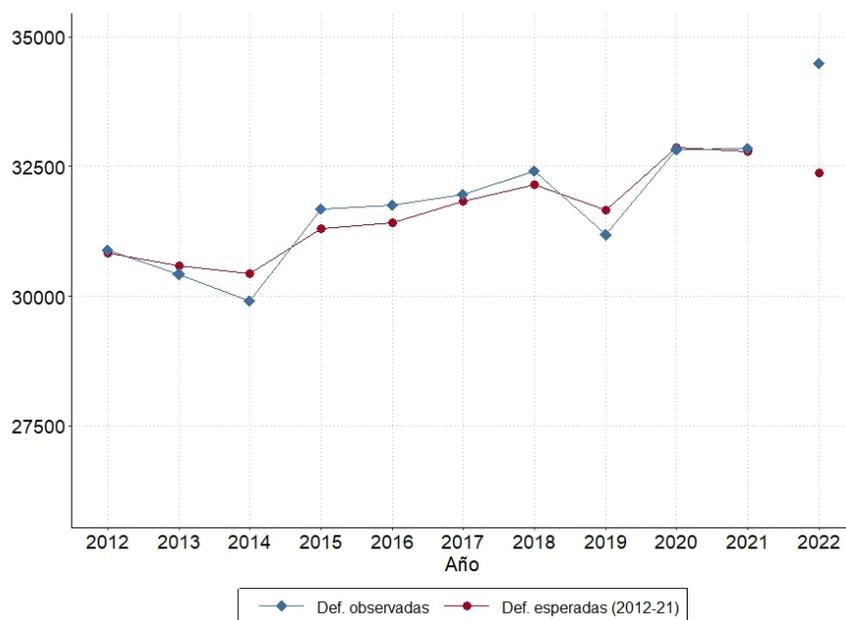


Figura 3.15: Evolución de la mortalidad esperada (a partir de los datos del período 2012-2021) y observada en Galicia desde el año 2012 al 2022. En el eje *X* están representados los años y en el eje *Y* el número de defunciones.

Para examinar con más detalle el exceso en la predicción del número de defunciones en el año 2022, se han agrupado por el capítulo de la causa de muerte. En la [Tabla 3.19](#), están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura 3.16](#).

Como ya se explicó anteriormente, se identificó un exceso de 2111.7 defunciones para el total anual. En la [Tabla 3.19](#), se observa que los capítulos que más contribuyen a esta infraestimación son: el capítulo *I* (enfermedades infecciosas y parasitarias) con un exceso de 975.3 defunciones, el capítulo *IV* (enfermedades endocrinas/nutricionales/metabólicas) con un exceso de 263.2, el capítulo *X* (enfermedades del sistema respiratorio) con un exceso de 363.7 y el capítulo *XVIII* (síntomas/signos/resultados anormales, no clasificados bajo otro concepto) con un exceso de 798.4. La situación del capítulo *I* se debe a que, aunque el COVID-19 ya existía, solo se disponía de registros de los años 2020 y 2021. Por otro lado, el exceso observado en el capítulo *XVIII* persiste debido a que, a pesar de la mayor experiencia del personal codificador, este capítulo sigue siendo más complejo ya que incluye las causas consideradas mal definidas. También cabe destacar que en el capítulo *IX* (enfermedades del sistema circulatorio) se sobreestima, obteniendo así un defecto de 452.1 defunciones en este capítulo.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	2291	1315.7	975.3	42.57 %
II	8324	8463.6	-139.6	-1.68 %
III	159	155.3	3.7	2.33 %
IV	1157	893.8	263.2	22.75 %
V	2480	2452.5	27.5	1.11 %
VI-VIII	1638	1723.7	-85.7	-5.23 %
IX	9159	9611.1	-452.1	-4.94 %
X	3508	3144.3	363.7	10.37 %
XI	1585	1521.7	63.3	3.99 %
XII	159	159.8	-0.8	-0.50 %
XIII	280	195.6	84.4	30.14 %
XIV	1236	1065.9	170.1	13.76 %
XV-XVII	63	63.6	-0.6	-0.95 %
XVIII	1034	235.6	798.4	77.21 %
XX	1414	1373.2	40.8	2.89 %
Total	34487	32375.3	2111.7	6.12 %

Tabla 3.19: Defunciones observadas, esperadas (a partir de los datos del período 2012-2021), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2022.

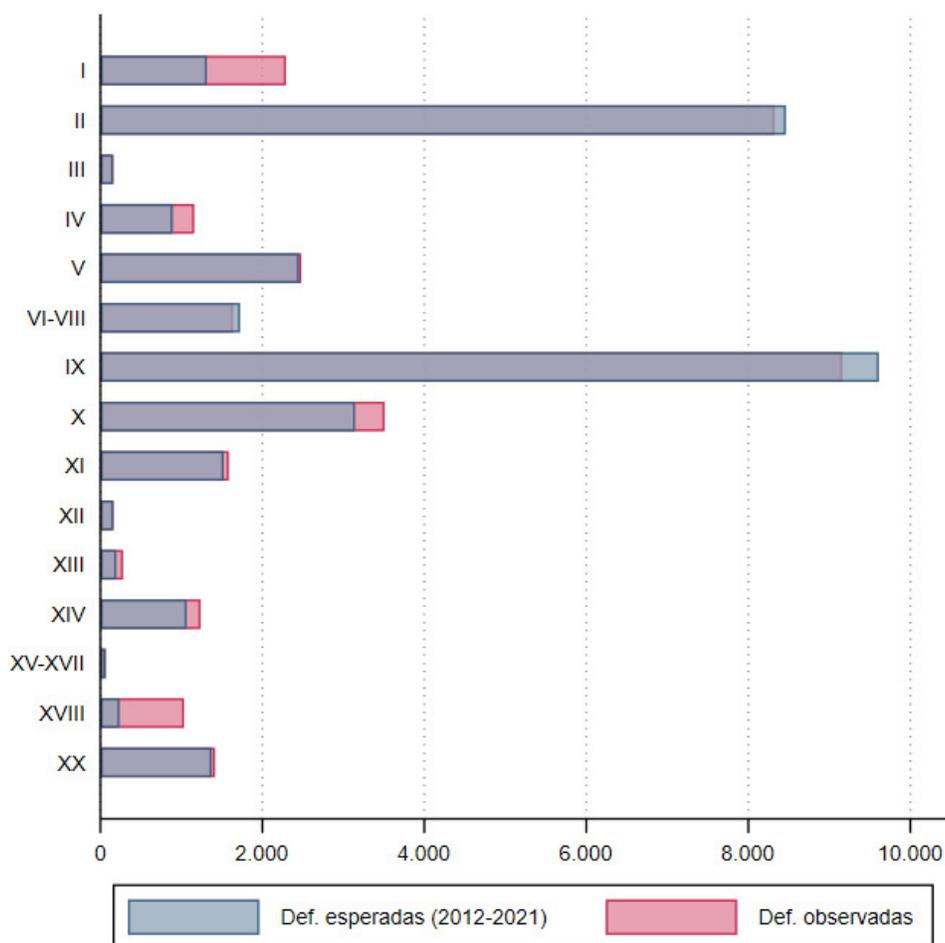


Figura 3.16: Defunciones observadas y esperadas, a partir de los datos del período 2012-2021 en función del capítulo. Galicia, 2022.

Para analizar con mayor profundidad lo que ocurre con cada causa de muerte, en la [Tabla 3.20](#) para las 36 causas seleccionadas están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas para el año 2022. Estos valores se presentan gráficamente en la [Figura 3.17](#).

En la [Tabla 3.20](#), se observa que las dos causas del capítulo *IV* (enfermedades endocrinas, nutricionales y metabólicas) contribuyen a la infraestimación con 144.3 defunciones la causa *IV:044* (diabetes mellitus) y 118.9 la causa *IV:Resto* (resto de causas del capítulo enfermedades endocrinas, nutricionales y metabólicas). Para el capítulo *X* (enfermedades del sistema respiratorio), las causas que más contribuyen al exceso de defunciones son la *X:063* (neumonía) y la *X:067* (otras enfermedades del sistema respiratorio), con un porcentaje de exceso de 23.44% y 8.80% sobre las 619 y 1885 defunciones observadas, respectivamente. Por otro lado, para el capítulo *IX* (enfermedades del sistema circulatorio), las causas que más contribuyen al defecto de defunciones son la *IX:054* (enfermedades hipertensivas), la *IX:056* (otras enfermedades isquémicas del corazón) y la *IX:058* (otras enfermedades del corazón), con un defecto de 102.4, 132.6 y 172.1 defunciones, respectivamente.

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	2291	1315.7	975.3	42.57 %
II:011	381	407.2	-26.2	-6.88 %
II:012	793	814.1	-21.1	-2.66 %
II:015	533	549.4	-16.4	-3.08 %
II:018	1547	1551	-4	-0.26 %
II:028	462	480.3	-18.3	-3.96 %
II:035	409	454.1	-45.1	-11.03 %
II:Resto	4199	4207.6	-8.6	-0.20 %
III	159	155.3	3.7	2.33 %
IV:044	808	663.7	144.3	17.86 %
IV:Resto	349	230.1	118.9	34.07 %
V:046	2401	2371.1	29.9	1.25 %
V:Resto	79	81.4	-2.4	-3.04 %
VI-VIII:050/052	963	999.1	-36.1	-3.75 %
VI-VIII:051	675	724.6	-49.6	-7.35 %
IX:054	1109	1211.4	-102.4	-9.23 %
IX:055	751	782.9	-31.9	-4.25 %
IX:056	1125	1257.6	-132.6	-11.79 %
IX:057:I50.9	1468	1470.5	-2.5	-0.17 %
IX:057:Resto	318	322	-4	-1.26 %
IX:058	1917	2089.1	-172.1	-8.98 %
IX:059	1935	1985	-50	-2.58 %
IX:Resto	536	492.5	43.5	8.12 %
X:063	619	473.9	145.1	23.44 %
X:064	774	774.6	-0.6	-0.08 %
X:066:J96.0.9	148	67.4	80.6	54.46 %
X:067	1885	1719.2	165.8	8.80 %
X:Resto	82	109.2	-27.2	-33.17 %
XI:072	1002	944.5	57.5	5.74 %
XI:Resto	583	577.2	5.8	0.99 %
XII	159	159.8	-0.8	-0.50 %
XIII	280	195.6	84.4	30.14 %
XIV	1236	1065.9	170.1	13.76 %
XV-XVII	63	63.6	-0.6	-0.95 %
XVIII	1034	235.6	798.4	77.21 %
XX	1414	1373.2	40.8	2.89 %
Total	34487	32375.3	2111.7	6.12 %

Tabla 3.20: Defunciones observadas, esperadas (a partir de los datos del período 2012-2021), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2022.

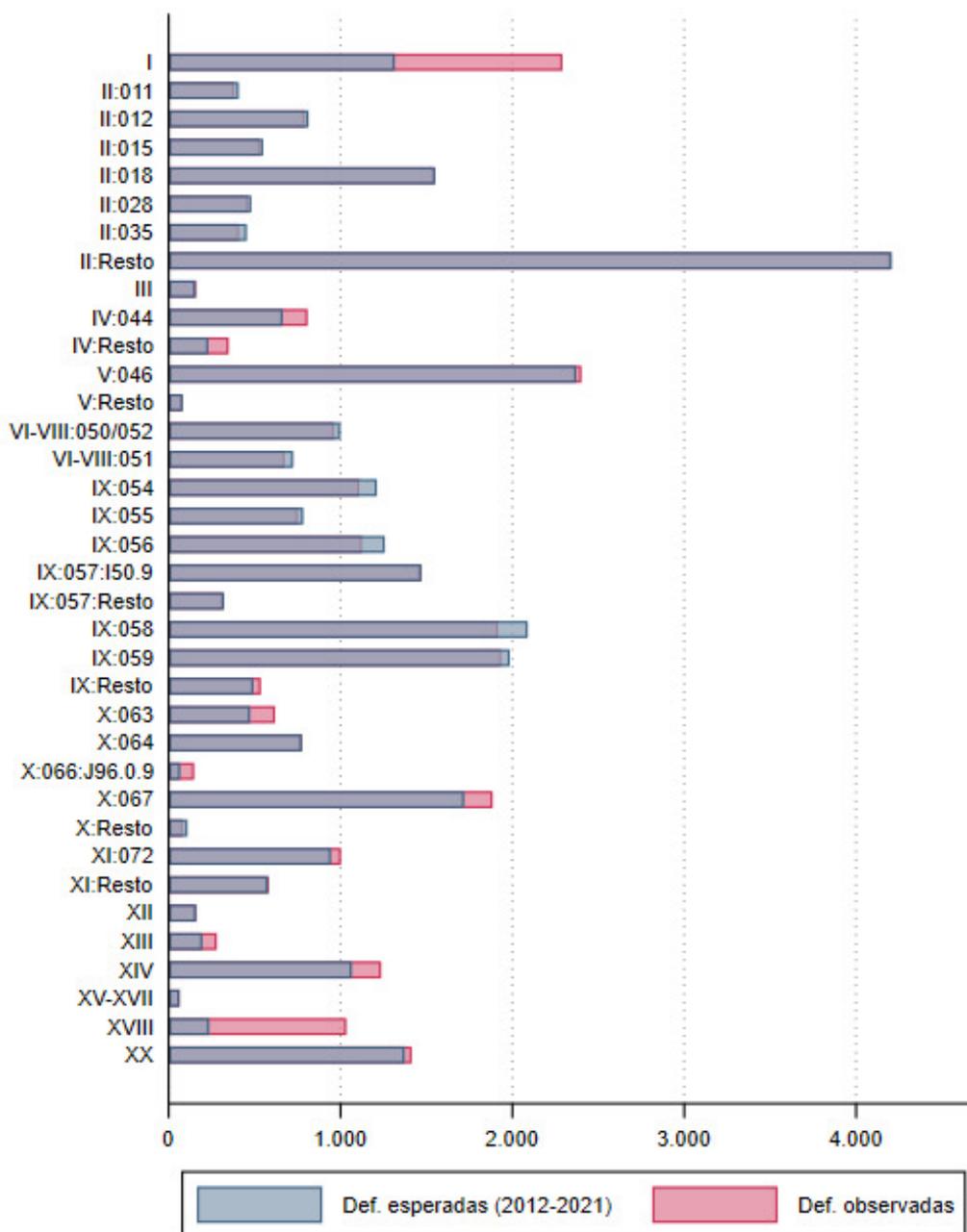


Figura 3.17: Defunciones observadas y esperadas, a partir de los datos del período 2012-2021 en función de la de las causas de muerte seleccionadas. Galicia, 2022.

Capítulo 4

Conclusiones y limitaciones

En este capítulo, se presentan las conclusiones obtenidas a partir de los resultados de aplicar el modelo mixto a los datos a estudio en el [Capítulo 3](#). Además, se lleva a cabo la comparación entre estos resultados y los explicados inicialmente en el [Capítulo 1](#).

Para el año 2019, se identificó un defecto de 1380.6 muertes observadas en comparación con las que cabría esperar tras ajustar los modelos mixtos, 31184 observadas frente a 32564.6 esperadas. Este defecto se justificó principalmente con el capítulo de enfermedades del sistema circulatorio, siendo el fallo cardíaco agudo la causa con mayor defecto de defunciones. Relacionado con esta causa, que es una de las afecciones mal definidas fuera del capítulo de síntomas/signos/resultados anormales, no clasificados bajo otro concepto, este último capítulo es el que presentó un mayor exceso de defunciones.

En el año 2020, se observó el comportamiento contrario, es decir, más muertes observadas (32823) que esperadas (32006.1), presentando así un exceso de 816.9 muertes. En este año, el capítulo de enfermedades infecciosas y parasitarias registró más mortalidad de la esperada (exceso de 1623.8 defunciones) debido a la aparición de la nueva enfermedad COVID-19 que se incluye en este capítulo. Además, destacó el capítulo de enfermedades del sistema respiratorio por presentar un acusado defecto de defunciones, siendo otras enfermedades del sistema respiratorio la causa que tuvo un mayor defecto.

Para el año 2021, se identificó un exceso de 412.8 muertes, 32853 observadas frente a 32440.2 esperadas. Se obtuvieron resultados muy parecidos a los del año anterior, ya que el capítulo de enfermedades infecciosas y parasitarias volvió a destacar con un exceso de 967.6 defunciones. Por otro lado, el capítulo de enfermedades del sistema respiratorio aumentó su defecto de defunciones, siendo nuevamente otras enfermedades del sistema respiratorio la causa con el mayor defecto.

Por último, para el año 2022, al igual que en los dos años anteriores, se identificó un exceso de 2111.7 muertes, con 34487 observadas frente a 32375.3 esperadas. En este año, el capítulo de enfermedades infecciosas y parasitarias y el de síntomas/signos/resultados anormales, no clasificados bajo otro concepto presentaron un acusado exceso de defunciones, con 975.3 y 798.4 muertes, respectivamente. Es importante destacar que el capítulo de enfermedades del sistema respiratorio presentó un exceso de 263.2 en vez de un defecto como sucedía en los dos años anteriores. En contraste, destacó el capítulo de enfermedades del sistema circulatorio por presentar un gran defecto de defunciones.

En los cuatro años a estudio, la mortalidad observada del capítulo *XVIII*, de causas mal definidas, fue superior a lo que cabría esperar destacando en mayor medida el año 2022.

Se observa una variación en los totales anuales entre los resultados introducidos inicialmente y los obtenidos mediante el modelo mixto. Para el año 2020, las 33643 defunciones esperadas presentadas inicialmente se ajustaron, tras la aplicación del modelo mixto, a 32006.1 defunciones esperadas, lo que las acerca más a las 32823 observadas. En el caso del año 2021, las 34235 defunciones esperadas se redujeron a 32440.2 tras la aplicación del modelo mixto, acercándose más a las 32853 observadas. Por último, para el año 2022, las defunciones esperadas pasaron de 34819 a 32375.3, alejándose aún más de las 34487 defunciones observadas. Al comparar los resultados del modelo mixto con los del análisis previo realizado en la DXSP, es importante tener en cuenta que este último se basó en un histórico de

cinco años. Es relevante destacar que los datos de 2019, en el estudio de la DXSP, mostraron una ligera disminución en comparación con años anteriores, lo que podría provocar una posible infraestimación en los resultados.

En cuanto a las limitaciones, aunque el modelo mixto se ajusta causa a causa y resulta efectivo en la predicción de defunciones a nivel individual por cada causa específica, al considerar el conjunto del año, las predicciones anuales pueden alejarse de lo esperado debido a los desvíos individuales por infraestimación o sobreestimación en cada causa.

En lo que respecta a las líneas futuras de investigación, se sugiere explorar el uso de datos funcionales como una posible extensión de este estudio, lo que podría proporcionar un nuevo enfoque. Además, se podría considerar la incorporación de variables de interés adicionales para mejorar la capacidad predictiva de los modelos utilizados.

Apéndice A

MoMo y EuroMOMO

En el año 2004, se creó el sistema de monitorización de la mortalidad diaria por todas las causas, MoMo¹, en el marco del “Plan de acciones preventivas contra los efectos de las temperaturas excesivas”. Este sistema de monitorización, coordinado por el Ministerio de Sanidad, se ha implementado en la Unidad de Vigilancia de la Mortalidad diaria del Centro Nacional de Epidemiología (CNE) del Instituto de Salud Carlos III.

Uno de los objetivos de MoMo es identificar las desviaciones de la mortalidad diaria observada con respecto a la esperada según las series históricas de mortalidad, lo que permite estimar de forma indirecta el impacto de cualquier evento de importancia en Salud Pública.

MoMo proporciona estimaciones de excesos de mortalidad por todas las causas y atribuibles a un exceso o defecto de la temperatura. Se calcula diariamente, por ámbito poblacional (nacional, de comunidades autónomas y provincial), sexo y grupo de edad (todas las edades, 0-14, 15-44, 45-64, 65-74, 75-84, 65 años o más y 85 años o más).

La [Figura A.1](#) es una captura de pantalla del Panel MoMo. En particular, se ven los resultados para Galicia en el período 2015-2022. En el gráfico, la línea gris representa la mortalidad observada y la línea azul la estimada. Además, debajo del gráfico, se encuentra una tabla con los datos que son los mismos que los recogidos en la [Tabla 1.1](#).

La información que utiliza MoMo para llevar a cabo las estimaciones procede de las siguientes fuentes de datos:

- Las defunciones diarias por todas las causas de los últimos diez años, sin incluir aquellas del año 2020 (por su comportamiento debido a la pandemia de COVID-19). Esta serie se obtiene de los datos consolidados del Instituto Nacional de Estadística (INE), hasta la fecha en la que están disponibles, y a partir de entonces del Ministerio de Justicia, que proporciona defunciones diarias por todas las causas procedentes de los registros civiles informatizados de municipios que corresponden a un 93 % de la población española.
- Las temperaturas según la Agencia Estatal de Meteorología (AEMET) a nivel provincial en el mismo período de tiempo, incluyendo el año en curso.
- La población por grupo de edad, sexo y provincia, extraídos del INE.

¹Sistema de Monitorización de la Mortalidad (MoMo): https://momo.isciii.es/panel_momo/

Región
Galicia

Sexo
Todos

Grupo de edad
Todos

Fecha
01/01/2015 - 31/12/2022

262.527

Observadas

249.430

Estimadas base

13.096

Exceso por todas las causas

2.046

Atribuibles a temperatura

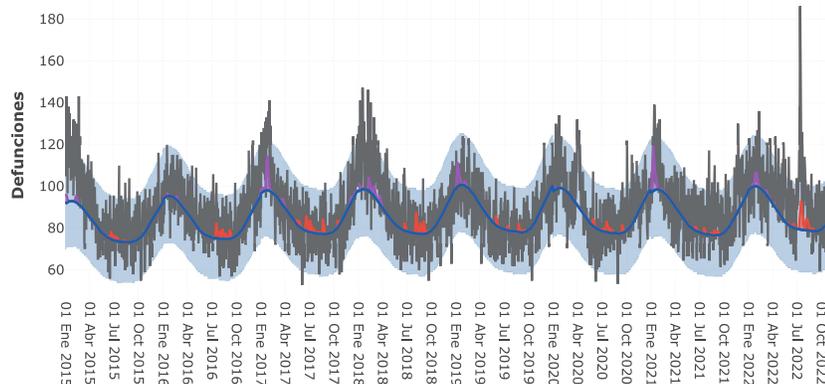
Mortalidad notificada, observada, esperada y atribuible a temperatura

Figura de mortalidad notificada y observada por todas las causas, estimadas base con un intervalo de confianza del 1% al 99% y defunciones atribuibles al exceso o defecto de temperatura, para el ámbito territorial, sexo, grupo de edad y rango de fechas seleccionado. Tanto la figura como la tabla muestra los resultados según los filtros seleccionados; hacer zoom en la gráfica no modificará estos resultados.

Defunciones observadas y estimadas, exceso de defunciones por todas las causas, defunciones atribuibles al exceso o defecto de temperatura y defunciones notificadasAgrupación: Día Semana Mes Año[Descargar CSV](#)

Fecha	Notificadas	Observadas	Estimadas base	Exceso por todas las causas	Atribuibles a temperatura
2022	35.164	35.164	31.814	3.350	356
2021	32.956	32.956	31.254	1.702	301
2020	33.042	33.042	31.626	1.416	145
2019	31.721	31.721	31.867	-146	203
2018	32.886	32.886	31.435	1.452	312
2017	32.450	32.450	31.289	1.161	417
2016	32.210	32.210	30.445	1.765	186
2015	32.098	32.098	29.701	2.397	126

Mortalidad notificada por todas las causas (aquella que se recibe realmente), mortalidad observada por todas las causas (aquella a la que se aplica una estimación por retraso para intentar aproximarnos a lo que se va a recibir), estimadas base, exceso por todas las causas (diferencia entre mortalidad observada y estimada base) y defunciones atribuibles al exceso o defecto temperatura (diferencia entre las defunciones esperadas con el efecto de la temperatura y las estimadas base), para el ámbito territorial, sexo, grupo de edad y rango de fechas seleccionado. Preste atención al rango de fechas y agrupación seleccionada, si su rango contiene una semana o mes incompleto, se mostrará también incompleto en la agregación.

Figura A.1: Captura de pantalla del Panel MoMo con los datos de Galicia en el período 2015-2022.

Como muchas amenazas a la salud pública no están restringidas por fronteras, los enfoques unificados a nivel internacional son fundamentales para detectar y estimar la magnitud del exceso de defunciones, ya que la combinación de datos aumenta el poder para detectar cambios rápidamente. El seguimiento de la mortalidad debe ser continuo para detectar cuándo y dónde se produce un exceso de mortalidad.

EuroMOMO² es un Proyecto Europeo de Vigilancia de la Mortalidad, cuyo objetivo es diseñar, al igual que MoMo, un sistema rutinario de vigilancia de la mortalidad en salud pública destinado a detectar y medir, en tiempo real, el exceso de muertes relacionadas con la gripe y otras posibles amenazas a la salud pública en los países europeos participantes.

Este objetivo se planteó originalmente en un proyecto apoyado por la DG Sanco que se desarrolló entre 2008 y 2011. Durante los tres años, múltiples socios de varios países europeos participaron en el desarrollo de un enfoque coordinado para hacer exactamente esto: monitoreo estandarizado de la mortalidad en tiempo real en toda Europa.

Las estadísticas oficiales de mortalidad nacional se proporcionan semanalmente desde los 28 países europeos o regiones subnacionales en la red colaborativa EuroMOMO, respaldada por el Centro Europeo para la Prevención y el Control de Enfermedades (ECDC) y la Organización Mundial de la Salud (OMS). En la mayoría de los países, se cubre toda la mortalidad de la población nacional. Sin embargo, en unos pocos países, solo se incluye una proporción de la mortalidad del país. Por ejemplo, en España se cubre el 93 % de la población, como se vio antes con MoMo.

La falta de una metodología común acordada para evaluar la mortalidad durante una crisis sanitaria importante que afecta a varios estados europeos limita el uso de esta fuente de información potencialmente muy poderosa. Por lo tanto, un mapeo oportuno del impacto de las amenazas a la salud sobre la mortalidad en diferentes países se beneficia enormemente de un enfoque uniforme.

La [Figura A.2](#) es una captura de pantalla de EuroMOMO. En particular, se ven los resultados de España, Portugal y Francia en el período 2019-2022.

²Proyecto Europeo de Monitorización de la Mortalidad (EuroMOMO): <https://www.euromomo.eu/>

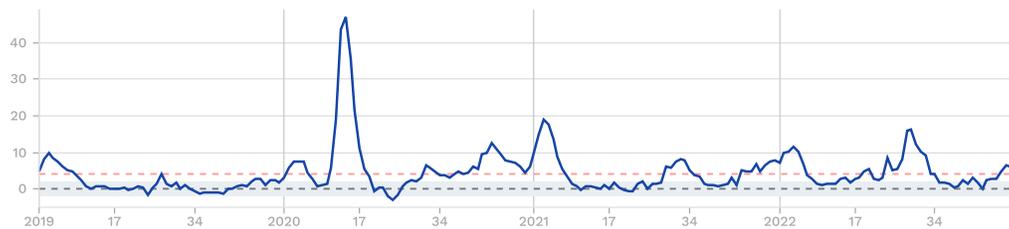
Z-scores by country

Graphs of the weekly z-score at the national level in the EuroMOMO partner countries and subnational regions.

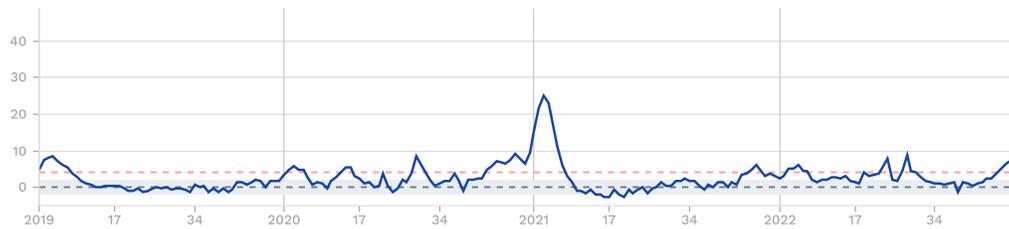
* Any use of this data in graphs or papers should cite "EuroMOMO (euromomo.eu) 2024". Filters applied to the graph will be applied to the downloaded data as well.

— Z-score - - - Baseline ■ Normal range - - - Substantial increase ■ Corrected for delay in registration

Spain



Portugal



France

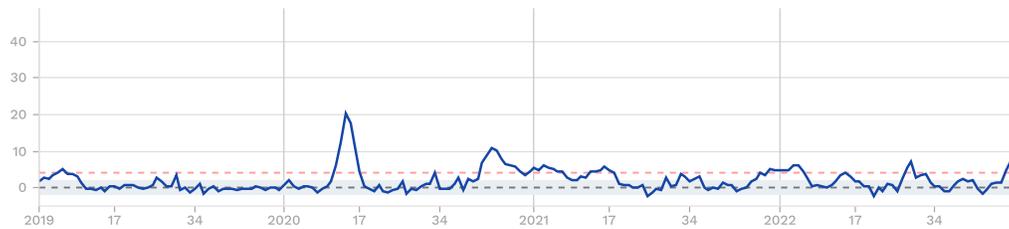


Figura A.2: Captura de pantalla de EuroMOMO con los datos de España, Portugal y Francia en el período 2019-2022.

Apéndice B

Tasa bruta, específica y ajustada de mortalidad

La **tasa bruta de mortalidad** es la proporción de personas que fallecen respecto al total de la población. Se calcula dividiendo el número total de defunciones entre la población total, ambas cantidades referidas a un área y período temporal concreto (se suele multiplicar por 100.000). Esta tasa bruta no puede usarse para comparar varias poblaciones o una misma población en distintos años puesto que no tiene en cuenta la estructura etaria de la población. Es decir, que una población presente una tasa bruta de mortalidad superior no quiere decir que tenga un problema de salud mayor, porque esta diferencia entre las tasas podría deberse a la diferencia en las estructuras de edad de las poblaciones que se están comparando. La **tasa específica de mortalidad** se obtiene como una tasa bruta pero referida a un grupo de edad concreto.

Para poder comparar poblaciones con diferentes estructuras etáreas existen diversos métodos de ajuste de tasas. La **tasa ajustada de mortalidad por el método directo** es una medida de resumen de las tasas específicas en los diferentes estratos de la población, habitualmente grupos de edad quinquenales. Se construye como una media ponderada de estas tasas específicas donde las ponderaciones proceden de una población de referencia llamada estándar. Las tasas ajustadas no tienen valor intrínseco, se usan únicamente para compararlas con otras obtenidas bajo las mismas condiciones. La fórmula de la tasa ajustada de mortalidad por el método directo viene dada por:

$$t_{\text{est}} = \sum_{j=1}^m \frac{p_{ej}}{p_e} t_j$$

donde

- m es el número de grupos de edad para el ajuste.
- c_j es el número de casos en el j -ésimo grupo de edad con $j = 1, \dots, m$.
- p_j es la población del j -ésimo grupo de edad con $j = 1, \dots, m$.
- p_{ej} es la población estándar del j -ésimo grupo de edad con $j = 1, \dots, m$.
- $p_e = \sum_{j=1}^m p_{ej}$ es la población estándar total.
- $t_j = \frac{c_j}{p_j}$ es la tasa específica del j -ésimo grupo de edad con $j = 1, \dots, m$.

Apéndice C

Lista reducida de causas de muerte CIE-10 y su correspondencia con la CIE-9

<i>Grupos de causas (desde 2016)^p</i>		<i>Código de la lista detallada CIE-10</i>	<i>Código de la lista detallada CIE-9</i>
001-102	Todas las causas	A00-Y89	001-E999
I. 001-008	Enfermedades infecciosas y parasitarias ¹	A00-B99, R75, U04.9, U07.0.1.2, U10.9	001-139, 279.5.6, 795.8
	00A COVID-19. Virus identificado	U07.1	
	00B COVID-19. Virus no identificado	U07.2	
	00C Síndrome inflamatorio multisistémico asociado con COVID-19, no especificado	U10.9	
	001 Enfermedades infecciosas intestinales	A00-A09	001-009
	002 Tuberculosis y sus efectos tardíos	A15-A19, B90	010-018, 137
	003 Enfermedad meningocócica	A39	036
	004 Septicemia	A40, A41	038
	005 Hepatitis vírica	B15-B19	070
	006 SIDA	B20-B24	279.5.6
	007 VIH+ (portador, evidencias de laboratorio del VIH,...)	R75	795.8
	008 Resto de enfermedades infecciosas y parasitarias y sus efectos tardíos	Resto A00-B99, U04.9,U07.0	Resto 001-139
II. 009-041	Tumores	C00-D48	140-239, 273.1.3, 289.8
	009 Tumor maligno del labio, de la cavidad bucal y de la faringe	C00-C14	140-149
	010 Tumor maligno del esófago	C15	150
	011 Tumor maligno del estómago	C16	151
	012 Tumor maligno del colon	C18	153
	013 Tumor maligno del recto, de la porción rectosigmoide y del ano	C19-C21	154
	014 Tumor maligno del hígado y vías biliares intrahepáticas	C22	155
	015 Tumor maligno del páncreas	C25	157
	016 Otros tumores malignos digestivos	Resto C15-C26, C45.1, C48	Resto 150-159
	017 Tumor maligno de la laringe	C32	161
	018 Tumor maligno de la tráquea, de los bronquios y del pulmón	C33, C34	162
	019 Otros tumores malignos respiratorios e intratorácicos	Resto C30-C39, C45.0.2	Resto 160-165
	020 Tumores malignos del hueso y de los cartílagos articulares	C40, C41	170
	021 Melanoma maligno de la piel	C43	172
	022 Otros tumores malignos de la piel y de los tejidos blandos	C44-C47, C49 (excepto C45.0.1.2)	171, 173
	023 Tumor maligno de la mama	C50	174,175
	024 Tumor maligno del cuello del útero	C53	180
	025 Tumor maligno de otras partes del útero	C54, C55	179,182
	026 Tumor maligno del ovario	C56	183.0
	027 Tumores malignos de otros órganos genitales femeninos	Resto C51-C58	Resto 179-184
	028 Tumor maligno de la próstata	C61	185
	029 Tumores malignos de otros órganos genitales masculinos	Resto C60-C63	186,187
	030 Tumor maligno del riñón, excepto pelvis renal	C64	189.0
	031 Tumor maligno de la vejiga	C67	188
	032 Otros tumores malignos de las vías urinarias	Resto C64-C68	Resto 188-189
	033 Tumor maligno del encéfalo	C71	191
	034 Otros tumores malignos neurológicos y endocrinos	Resto C69-C75	Resto 190-194
	035 Tumor maligno de sitios mal definidos, secundarios y de sitios no especificados	C76, C80	195,199
	036 Tumores malignos del tejido linfático, de los órganos hematopoyéticos y de tejidos afines (excepto leucemia)	C81-C90, C96	200-203, 273.3
	037 Leucemia	C91-C95	204-208
	038 Tumores in situ	D00-D09	230-234
	039 Tumores benignos	D10-D36	210-229
	040 Síndrome mielodisplásico ²	D46	289.8
	041 Otros tumores de comportamiento incierto o desconocido	D37-D45, D47, D48	235-239, 273.1
III. 042-043	Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan al mecanismo de la inmunidad	D50-D89	273.0.2, 279-289 (excepto 279.5.6, 289.8)
	042 Enfermedades de la sangre y de los órganos hematopoyéticos	D50-D77	280-289 (excepto 289.8)
	043 Ciertos trastornos que afectan al mecanismo de la inmunidad	D80-D89	273.0.2, 279 (excepto 279.5.6)
IV. 044-045	Enfermedades endocrinas, nutricionales y metabólicas	E00-E90	240-278, 330.0.1 (excepto 273.0.1.2.3, 274)
	044 Diabetes mellitus	E10-E14	250
	045 Otras enfermedades endocrinas, nutricionales y metabólicas	Resto E00-E90	Resto 240-278, 330.0.1 (excepto 273.0.1.2.3, 274)
<i>Grupos de causas (desde 2016)^p</i>		<i>Código de la lista detallada CIE-10</i>	<i>Código de la lista detallada CIE-9</i>
V. 046-049	Trastornos mentales y del comportamiento	F00-F99	290-319

	<i>Grupos de causas (desde 2016)⁰</i>	<i>Código de la lista detallada CIE-10</i>	<i>Código de la lista detallada CIE-9</i>
	046 Trastornos mentales orgánicos, senil y presenil	F00-F09	290
	047 Trastornos mentales debidos al uso de alcohol	F10	291, 303
	048 Trastornos mentales debidos al uso de drogas (drogodependencia, toxicomanía)	F11-F16, F18, F19	304, 305
	049 Otros trastornos mentales y del comportamiento	Resto F00-F99	Resto 290-319
VI-VIII. 050-052	Enfermedades del sistema nervioso y de los órganos de los sentidos	G00-H95	320-389, 435 (excepto 330.0.1)
	050 Meningitis (otras en 003)	G00-G03	320-322
	051 Enfermedad de Alzheimer	G30	331.0
	052 Otras enfermedades del sistema nervioso y de los órganos de los sentidos	Resto de G00-H95	Resto 320-389, 435 (excepto 330.0.1)
IX. 053-061	Enfermedades del sistema circulatorio³	I00-I99	390-459, (excepto 427.5, 435, 446, 459.0)
	053 Enfermedades cardíacas reumáticas crónicas	I05-I09	393-398
	054 Enfermedades hipertensivas	I10-I15	401-405
	055 Infarto agudo de miocardio	I21	410
	056 Otras enfermedades isquémicas del corazón	I20, I24-I25	411-414
	057 Insuficiencia cardíaca	I50	428
	058 Otras enfermedades del corazón	I00-I02, I26-I49, I51	390-392, 415-417, 420-427, 429 (excepto 427.5)
	059 Enfermedades cerebrovasculares	I60-I69	430-434, 436-438
	060 Aterosclerosis	I70	440
	061 Otras enfermedades de los vasos sanguíneos	I71-I99	441-459 (excepto 446, 459.0)
X. 062-067	Enfermedades del sistema respiratorio	J00-J99	460-519, 786.0
	062 Influenza (gripe) (incluye gripe por virus de la influenza pandémica o zoonótica identificados)	J09-J11	487
	063 Neumonía	J12-J18	480-486
	064 Enfermedades crónicas de las vías respiratorias inferiores (excepto asma)	J40-J44, J47	490-492, 494-496
	065 Asma	J45, J46	493
	066 Insuficiencia respiratoria ⁴	J96	786.0
	067 Otras enfermedades del sistema respiratorio	Resto J00-J99	Resto 460-519
XI. 068-072	Enfermedades del sistema digestivo	K00-K93	520-579
	068 Úlcera de estómago, duodeno y yeyuno	K25-K28	531-534
	069 Enteritis y colitis no infecciosas	K50-K52	555, 556, 558
	070 Enfermedad vascular intestinal	K55	557
	071 Cirrosis y otras enfermedades crónicas del hígado	K70, K72.1, K73, K74, K76.1.9	571
	072 Otras enfermedades del sistema digestivo	Resto K00-K93	Resto 520-579
XII. 073	Enfermedades de la piel y del tejido subcutáneo	L00-L99	680-709
XIII. 074-076	Enfermedades del sistema osteomuscular y del tejido conjuntivo	M00-M99	714, 715, 724, 446, 710-739
	074 Artritis reumatoide y osteoartritis	M05, M06, M15-M19	714, 715
	075 Osteoporosis y fractura patológica	M80-M82, M84.4	733
	076 Otras enfermedades del sistema osteomuscular y del tejido conjuntivo	Resto M00-M99	Resto 710-739, 274, 446
XIV. 077-080	Enfermedades del sistema genitourinario	N00-N99	580-629
	077 Enfermedades del riñón y del uréter	N00-N29	580-594
	078 Enfermedades de los órganos genitales masculinos	N40-N51	600-608
	079 Enfermedades de los órganos genitales femeninos y trastornos de la mama	N60-N64, N70-N98	610, 611, 614-629
	080 Otras enfermedades del sistema genitourinario	Resto N00-N99	Resto 580-629
XV. 081	Embarazo, parto y puerperio	O00-O99	630-676
XVI. 082	Afecciones originadas en el periodo perinatal	P00-P96	760-779
XVII. 083-085	Malformaciones congénitas, deformidades y anomalías cromosómicas	Q00-Q99	740-759
	083 Malformaciones congénitas del sistema nervioso	Q00-Q07	740-742
	084 Malformaciones congénitas del sistema circulatorio	Q20-Q28	745-747
	085 Otras malformaciones congénitas, deformidades y anomalías cromosómicas	Resto Q00-Q99	Resto 740-759
XVIII. 086-089	Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte¹	R00-R74, R76-R99	427.5, 459.0, 780-799 (excepto 786.0, 795.8, 795.8)
	086 Paro cardíaco, muerte sin asistencia y otra causa desconocida de mortalidad	R98, R99	427.5, 798.9, 799.9
	087 Senilidad	R54	797
	088 Muerte súbita infantil	R95	798.0
	089 Resto de síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte	Resto R00-R74, R76-R99	459.0, Resto 780-799 (excepto 786.0, 795.8)
XX. 090-102	Causas externas de mortalidad⁵	U12.9, V01-Y89	E800-E999
	090 Accidentes de tráfico	V01-V06.1.9 V09.2.3.9 V10-V18.3.4.5.9 V19.4.5.6.9	E810-E819, E826

Grupos de causas (desde 2016) ⁰	Código de la lista detallada CIE-10	Código de la lista detallada CIE-9
	V20-V28.3.4.5.9	
	V29.4.5.6.9	
	V30-V38.4.5.6.7.9	
	V39.4.5.6.9	
	V40-V48.4.5.6.7.9	
	V49.4.5.6.9	
	V50-V58.4.5.6.7.9	
	V59.4.5.6.9	
	V60-V68.4.5.6.7.9	
	V69.4.5.6.9	
	V70-V78.4.5.6.7.9	
	V79.4.5.6.9	
	V80.2.3.4.5	
	V81.1	
	V82.1.9	
	V83.0.1.2.3.4	
	V84.0.1.2.3.4	
	V85.0.1.2.3.4	
	V86.0.1.2.3.4	
	V87.0.1.2.3.4.5.6.7.8.9	
	V89.2.3.9	
091 Otros accidentes de transporte	Resto de V01-V99	E800-E807, E820-E825, E827-E848
092 Caídas accidentales	W00-W19	E880-E888 (excepto E887)
093 Ahogamiento, sumersión y sofocación accidentales	W65-W84	E910-E915
094 Accidentes por fuego, humo y sustancias calientes	X00-X19	E890-E899, E924
095 Envenenamiento accidental por psicofármacos y drogas de abuso	X41, X42, X44, X45	E850.0.9, E851-E855, E858.9, E860
096 Otros envenenamientos accidentales	Resto de X40-X49	Resto de E850-E869
097 Otros accidentes	Resto W00-X59	Resto E800-E849, Resto E880-E928
098 Suicidio y lesiones autoinfligidas	X60-X84	E950-E959
099 Agresiones (homicidio)	X85-Y09	E960-E969
100 Eventos de intención no determinada	Y10-Y34	E980-E989
101 Complicaciones de la atención médica y quirúrgica	Y40-Y84	E870-E879, E930-E949
102 Otras causas externas y sus efectos tardíos	U12.9, resto Y35-Y89	Resto E929-E999

(Colaboración Registros de mortalidad de las CCAA - INE)

⁰ La CIE es una clasificación viva y por tanto afectada periódicamente por ciertas modificaciones, aunque estas son hasta ahora despreciables; para facilitar su consulta, se indica el año de correspondencia y se marcan en **rojo-tachado**/ azul nuevo, las modificaciones respecto a la lista anterior. Más información en Metodología.

¹ Por motivos de comparabilidad y recomendación de la OPS, se incluyen en el capítulo I. Enfermedades infecciosas y parasitarias el VIH+ (R75) (795.8), el SARS - Síndrome Respiratorio Agudo Severo (U04.9), Trastorno relacionado con el vapeo (U07.0) y por tanto desaparecen de su clasificación original en el capítulo XVIII. Síntomas, signos y hallazgos anormales clínicos y de laboratorio (R75) y en el capítulo XXII. Códigos para propósitos especiales (U04.9). La OMS modifica para el año 2019 y siguientes el código de la enfermedad del virus Zika, no especificada, se elimina el U06.9 inicialmente propuesto y se crean el A92.5 y P35.4.

Dentro del capítulo I. Enfermedades infecciosas y parasitarias se incluyen también la COVID-19. Virus identificado (U07.1), COVID-19. Virus no identificado (U07.2) y Síndrome inflamatorio multisistémico asociado con COVID-19, no especificado (U10.9)

² Por motivos de comparabilidad, se desagrega el S. mielodisplásico (D46) (289.8) ya que la CIE-9 lo encuadraba en el capítulo IV. Enfermedades de la sangre y de los órganos hematopoyéticos.

³ Por motivos de comparabilidad, se excluye el Paro cardíaco (R98,R99) (427.5), ya que la CIE-10 lo encuadra en el capítulo XVIII. Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte.

⁴ Por motivos de comparabilidad con la CIE-10, se desagrega la Insuficiencia respiratoria (J96) (786.0) ya que la CIE-9 lo encuadraba en el capítulo XVI. Signos, síntomas y estados morbosos mal definidos.

⁵ Dentro del capítulo XX. Causas externas de mortalidad se incluye también las Vacunas COVID-19 que causan efectos adversos en uso

Apéndice D

Otros modelos

A partir de la conclusión de la [Sección 3.1](#), se emplearon los modelos mixtos, introducidos en la [Subsección 2.2.5](#), para ajustar los datos considerando que la variable `defunciones` sigue una distribución Binomial Negativa. Para cada año a predecir (2019-2022), se realizó el ajuste del modelo a partir del período que empieza en el año 1999 y termina en el año inmediatamente anterior al predicho.

Para el ajuste de los modelos, validaciones y predicciones se usó , en particular, se usó la librería `glmmTMB` que emplea la aproximación de Laplace para realizar el ajuste. Se utilizó  en lugar de Stata para ajustar los modelos mixtos con distribución Binomial Negativa debido a las limitaciones de Stata en este contexto. En concreto, Stata no proporciona la opción de incorporar una componente autorregresiva, `ar1`, a los modelos mixtos algo que se necesitará al hacer los ajustes como se verá posteriormente.

Empezando por el año 2019, para realizar el ajuste del modelo se tomó como período histórico de referencia el 1999-2018. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. Es decir, se ajustó un modelo para la variable respuesta `defunciones` considerando solo el intercepto y la población expuesta y, por otra parte se ajustó otro modelo considerando solo el intercepto y la población expuesta pero teniendo en cuenta los efectos aleatorios en función de la causa de defunción.

El valor del AIC para el modelo sin efectos aleatorios fue 170177.1 y el del modelo con efectos aleatorios fue 165543.4 ([Tabla D.1](#)). Por lo tanto, el modelo con menor AIC es el modelo con efectos aleatorios.

Ahora, a partir del modelo con efectos aleatorios se ajustaron modelos con intercepto aleatorio tomando como variables explicativas: el `sexo`, el `grupo de edad` y la `causa`, considerando las tres variables categóricas, también se tuvo en cuenta la población expuesta. Además, se ajustaron modelos con interacción `causa-sexo` y `causa-grupo de edad`.

Además de los modelos con intercepto aleatorio, se ajustaron modelos con intercepto y pendiente aleatorios. De estos modelos el único que convergió fue el modelo con:

- Variables explicativas: `sexo` y `grupo de edad`.
- Población expuesta
- Variable grupo: `causa`.
- Pendiente aleatoria: `sexo`.

Los demás modelos con intercepto y pendiente aleatorios no convergieron, ya que todos presentaron el mismo error relacionado con que la matriz hessiana no era definida positiva.

Como se indicará posteriormente, en algunos modelos se ha considerado una componente autorregresiva, `ar1`, para incluir la dependencia temporal. Para la selección del modelo se escoge el que tenga un menor AIC, los valores del AIC de todos los modelos están recogidos en la [Tabla D.1](#), en la cual también se recoge la información de como se hizo el ajuste de cada modelo.

Variables explicativas del modelo	AIC
<ul style="list-style-type: none"> ■ Población expuesta. 	170177.1
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variable grupo: <i>causa</i>. 	165543.4
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i> y <i>grupo de edad</i>. ■ Variable grupo: <i>causa</i>. ■ Componente <i>ar1</i>. 	126978.4
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i>, <i>grupo de edad</i> y <i>causa</i>. ■ Variable grupo: <i>causa</i>. ■ Componente <i>ar1</i>. 	126915
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i>, <i>grupo de edad</i>, <i>causa</i> e <i>interacción causa-sexo</i>. ■ Variable grupo: <i>causa</i>. ■ Componente <i>ar1</i>. 	124622.7
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i>, <i>grupo de edad</i>, <i>causa</i> e <i>interacción causa-grupo de edad</i>. ■ Variable grupo: <i>causa</i>. ■ Componente <i>ar1</i>. 	107645.4
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i>, <i>grupo de edad</i>, <i>causa</i>, <i>interacción causa-sexo</i> e <i>interacción causa-grupo de edad</i>. ■ Variable grupo: <i>causa</i>. ■ Componente <i>ar1</i>. 	99368.6
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: <i>sexo</i> y <i>grupo de edad</i>. ■ Variable grupo: <i>causa</i>. ■ Pendiente aleatoria: <i>sexo</i>. ■ Componente <i>ar1</i>. 	124871

Tabla D.1: Valores del AIC de los distintos modelos mixtos ajustados para el año 2019. Estos resultados se obtienen con la función AIC de .

Por tanto, se puede observar que el menor valor del AIC es 99368.6 que se corresponde con el modelo que tiene como variables explicativas: **sexo**, **grupo de edad**, **causa**, la interacción **causa-sexo** y la interacción **causa-grupo de edad**. Algunos de los coeficientes de este modelo están recogidos en la [Tabla D.2](#).

Coeficientes	Estimador	Error típico	Z-value	P-valor
Intercepto	-10.59	0.26	-40.21	< 0.01
gidade2	-1.82	0.48	-3.77	< 0.01
gidade3	-1.20	0.37	-3.28	< 0.01
⋮	⋮	⋮	⋮	⋮
sexo2	-0.63	0.03	-20.40	< 0.01
as.factor(causa)II:011	-5.93	3.83	-1.55	0.12
as.factor(causa)II:012	-6.44	3.80	-1.69	0.09
⋮	⋮	⋮	⋮	⋮
gidade2:as.factor(causa)II:011	-2.86	38.32	-0.07	0.94
gidade3:as.factor(causa)II:011	-5.21	87.28	-0.06	0.95
⋮	⋮	⋮	⋮	⋮
sexo2:as.factor(causa)II:011	-0.12	0.04	-2.85	< 0.01
sexo2:as.factor(causa)II:012	-0.002	0.04	-0.06	0.95
⋮	⋮	⋮	⋮	⋮

Tabla D.2: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2019. El modelo consta de un total de 722 coeficientes.

Para la validación del modelo se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados ([Figura D.1](#)) y el segundo, es el qq-plot de los residuos de la deviance ([Figura D.2](#)). En la [Figura D.1](#) se observa que existen patrones no lineales con respecto a los valores ajustados y en la [Figura D.2](#) se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

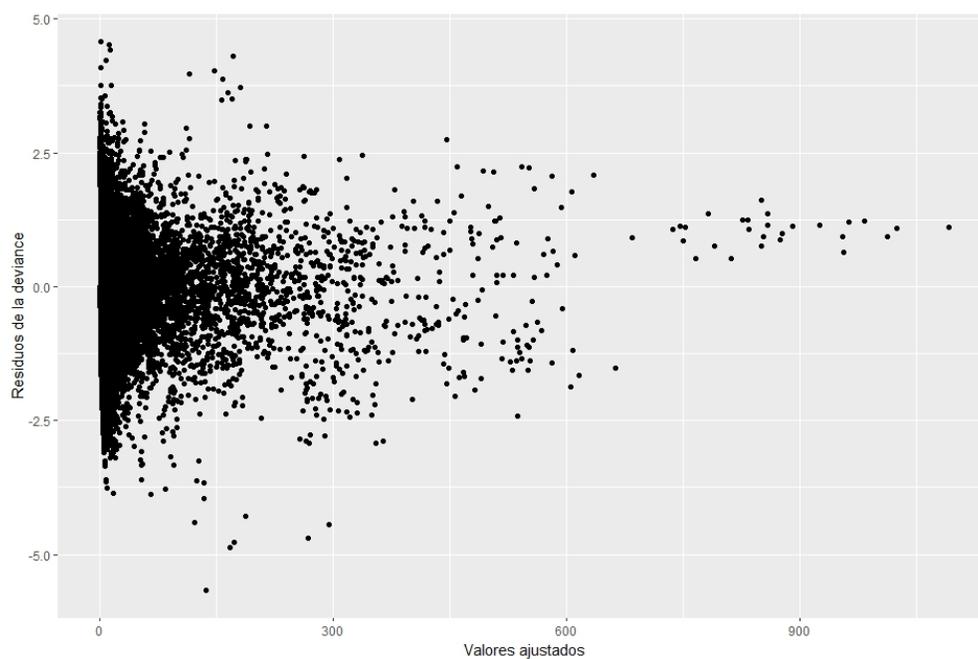


Figura D.1: Residuos de la deviance del modelo seleccionado para el año 2019 frente a los valores ajustados.

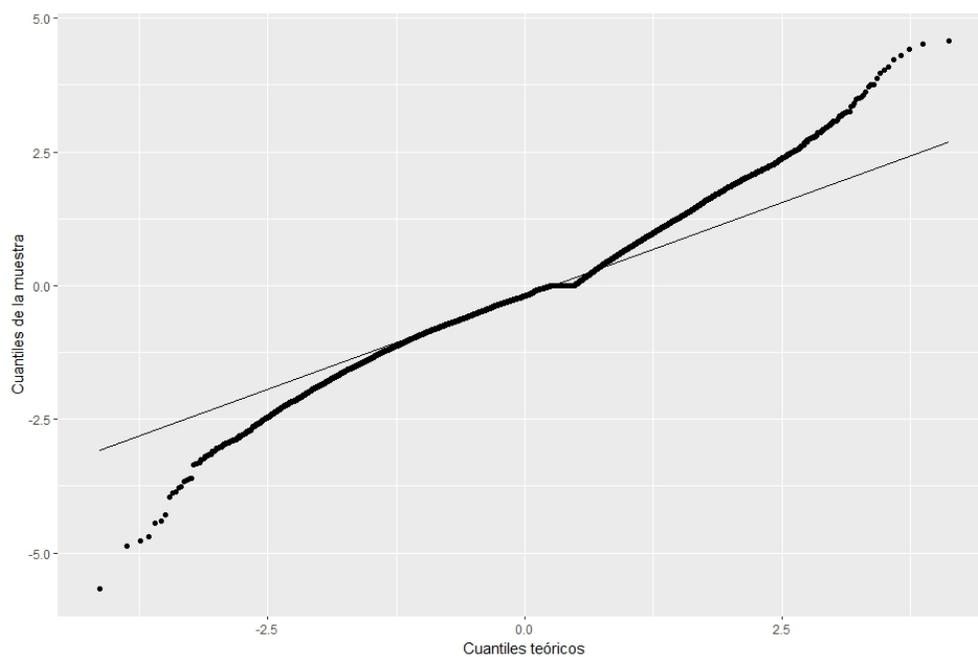


Figura D.2: QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2019.

En la [Tabla D.3](#) para cada causa seleccionada para el año 2019 están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.3](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	412	479.9	-67.9	-16.48 %
II:011	428	639.6	-211.6	-49.44 %
II:012	836	884.7	-48.7	-5.83 %
II:015	504	486.3	17.7	3.51 %
II:018	1500	1607.3	-107.3	-7.15 %
II:028	504	635.3	-131.3	-26.05 %
II:035	431	597.2	-166.2	-38.56 %
II:Resto	4142	4352.7	-210.7	-5.09 %
III	157	116.4	40.6	25.86 %
IV:044	471	732.1	-261.1	-55.44 %
IV:Resto	162	128.3	33.7	20.80 %
V:046	2414	1909.5	504.5	20.90 %
V:Resto	78	84.6	-6.6	-8.46 %
VI-VIII:051	798	664.8	133.2	16.69 %
VI-VIII:052	890	672.3	217.7	24.46 %
VI-VIII:Resto	5	9.4	-4.4	-88.00 %
IX:054	803	757.2	45.8	5.70 %
IX:055	869	1780.5	-911.5	-104.89 %
IX:056	1219	1606.8	-387.8	-31.81 %
IX:057:I50.9	1236	1680.6	-444.6	-35.97 %
IX:057:Resto	281	539.4	-258.4	-91.96 %
IX:058	2140	2109.3	30.7	1.43 %
IX:059	2142	3608.8	-1466.8	-68.48 %
IX:Resto	499	848.5	-349.5	-70.04 %
X:063	611	936.1	-325.1	-53.21 %
X:064	937	1840.4	-903.4	-96.41 %
X:066:J96.0	5	23	-18	-360 %
X:066:J96.9	26	86.4	-60.4	-232.31 %
X:067	1830	2111.2	-281.2	-15.37 %
X:Resto	131	214.2	-83.2	-63.51 %
XI:072	961	877.3	83.7	8.71 %
XI:Resto	598	761.6	-163.6	-27.36 %
XII	140	158.1	-18.1	-12.93 %
XIII	169	253.8	-84.8	-50.18 %
XIV	964	884.3	79.7	8.27 %
XV-XVII	64	94.8	-30.8	-48.12 %
XVIII	577	781.7	-204.7	-35.48 %
XX	1250	1470.9	-220.9	-17.67 %
Total	31184	37425	-6241	-20.01 %

Tabla D.3: Defunciones observadas, esperadas (a partir de los datos del período 1999-2018), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2019.

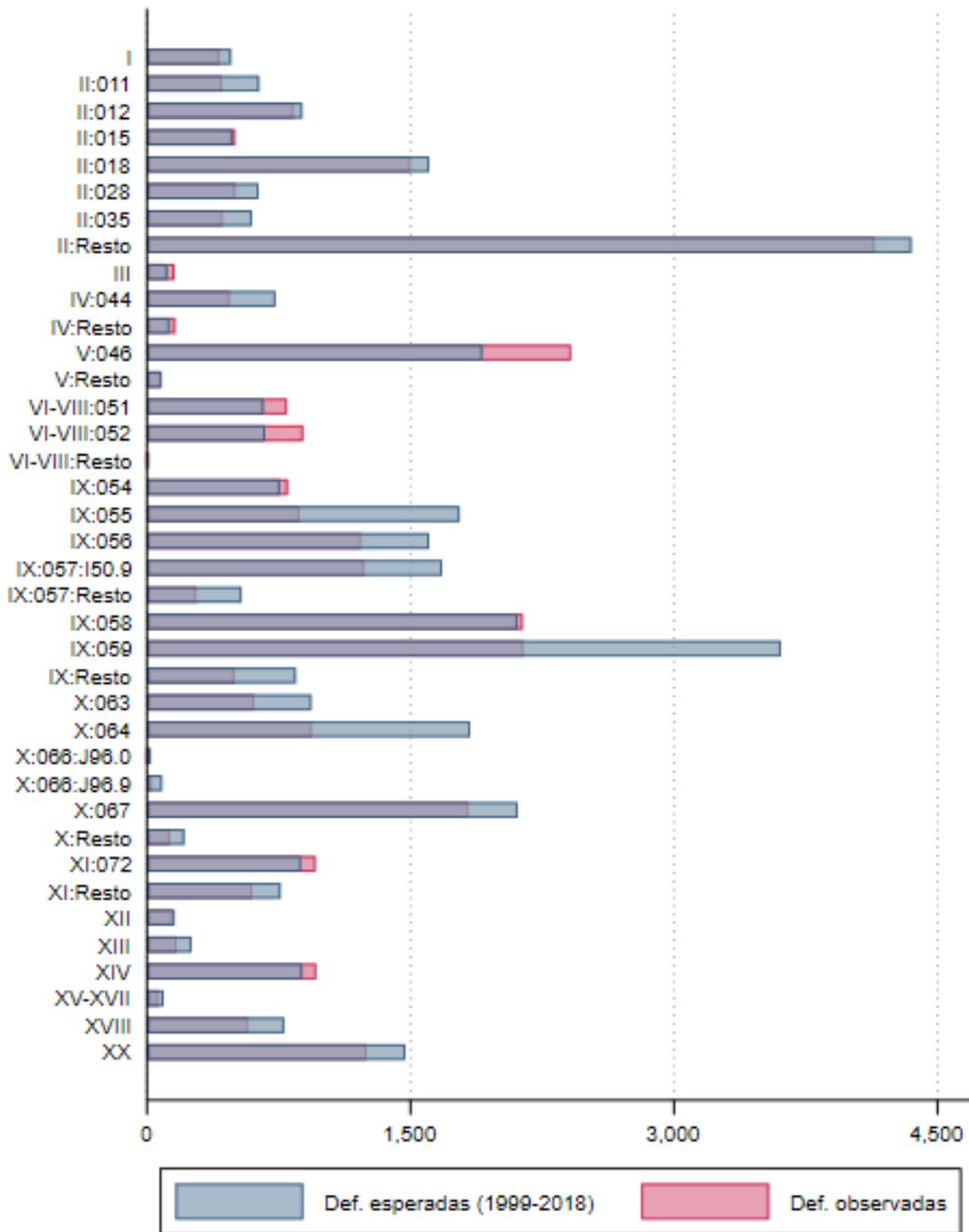


Figura D.3: Defunciones observadas y esperadas, a partir de los datos del período 1999-2018 en función de la de las causas de muerte seleccionadas. Galicia, 2019.

Agrupando por el capítulo de la causa de muerte, en la [Tabla D.4](#) están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.4](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	412	479.9	-67.9	-16.48 %
II	8345	9203.1	-858.1	-10.28 %
III	157	116.4	40.6	25.86 %
IV	633	860.4	-227.4	-35.92 %
V	2492	1994.1	497.9	19.98 %
VI-VIII	1693	1346.5	346.5	20.47 %
IX	9189	12931	-3742	-40.72 %
X	3540	5211.3	-1671.3	-47.21 %
XI	1559	1638.9	-79.9	-5.13 %
XII	140	158.1	-18.1	-12.93 %
XIII	169	253.8	-84.8	-50.18 %
XIV	964	884.3	79.7	8.27 %
XV-XVII	64	94.8	-30.8	-48.12 %
XVIII	577	781.7	-204.7	-35.48 %
XX	1250	1470.9	-220.9	-17.67 %
Total	31184	37425	-6241	-20.01 %

Tabla D.4: Defunciones observadas, esperadas (a partir de los datos del período 1999-2018), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2019.

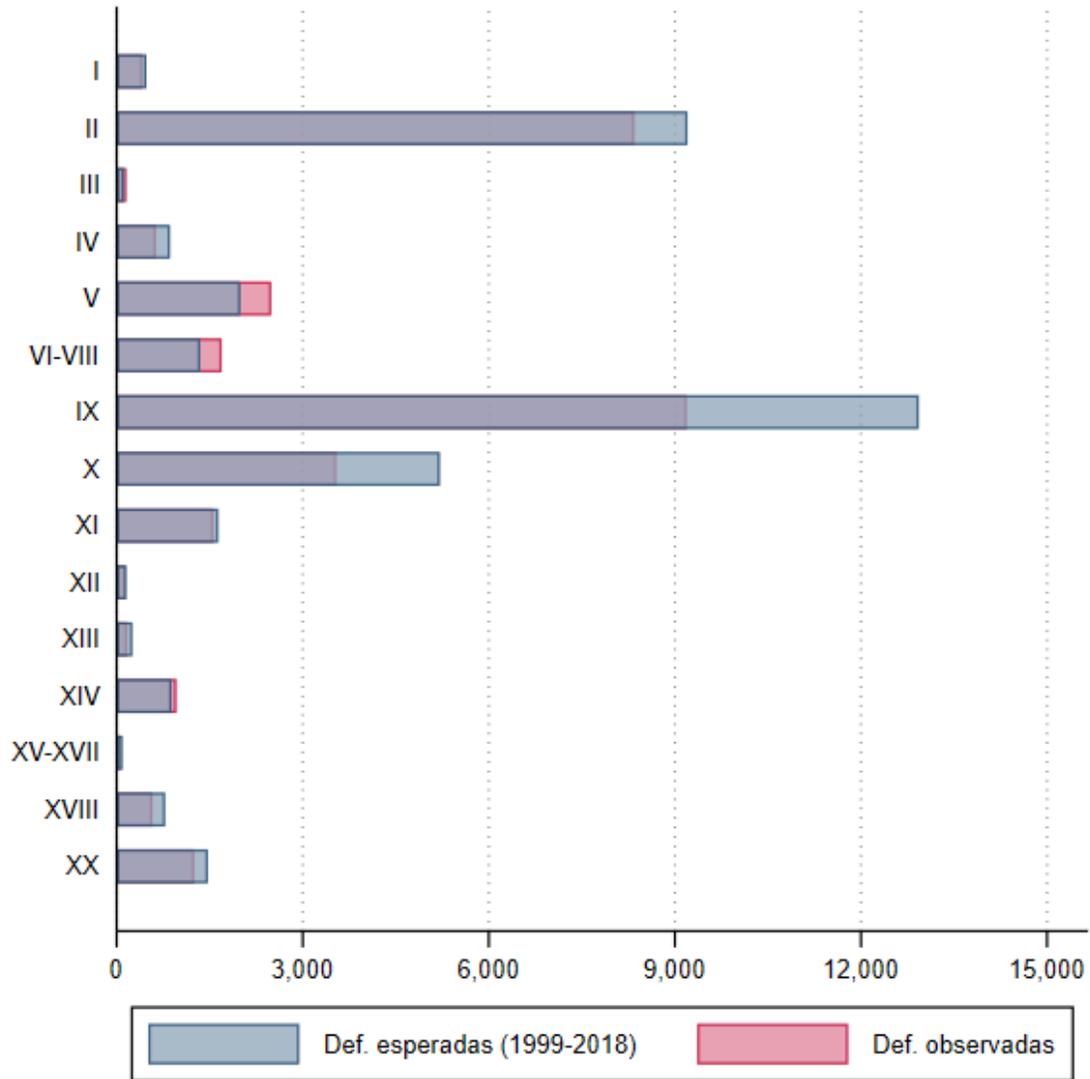


Figura D.4: Defunciones observadas y esperadas, a partir de los datos del período 1999-2018 en función del capítulo de la causa de muerte. Galicia, 2019.

A continuación, para el ajuste del modelo del año 2020 se siguió el mismo procedimiento que para el 2019. En este caso, se tomó como período histórico de referencia el 1999-2019. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. El valor del AIC para el modelo sin efectos aleatorios fue 178282.7 y el del modelo con efectos aleatorios fue 173412.6 (Tabla D.5). Por lo tanto, el modelo con menor AIC es el modelo con efectos aleatorios.

Variables explicativas del modelo	AIC
<ul style="list-style-type: none"> ■ Población expuesta. 	178282.7
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variable grupo: causa. 	173412.6
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	132831.1
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad y causa. ■ Variable grupo: causa. ■ Componente ar1. 	132767.8
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-sexo. ■ Variable grupo: causa. ■ Componente ar1. 	130356.8
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	112717.6
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa, interacción causa-sexo e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	104151.9
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Pendiente aleatoria: sexo. ■ Componente ar1. 	130607.3

Tabla D.5: Valores del AIC de los distintos modelos mixtos ajustados para el año 2020. Estos resultados se obtienen con la función AIC de .

Ahora, a partir del modelo con efectos aleatorios se ajustaron modelos con intercepto aleatorio tomando como variables explicativas las mismas que para el año 2019. También se ajustaron modelos con intercepto y pendiente aleatorios. De estos modelos, como ocurrió con el año 2019, el único que convergió fue el modelo con:

- Variables explicativas: **sexo** y **grupo de edad**.
- Población expuesta
- Variable grupo: **causa**.
- Pendiente aleatoria: **sexo**.

Los demás modelos con intercepto y pendiente aleatorios no convergieron, ya que todos presentaron el mismo error relacionado con que la matriz hessiana no era definida positiva. Como se indica en la [Tabla D.5](#), en algunos modelos se ha considerado una componente autorregresiva, **ar1**, para incluir la dependencia temporal.

Por tanto, en la [Tabla D.5](#) se puede observar que el menor valor del AIC es 104151.9 que se corresponde con el modelo que tiene como variables explicativas: **sexo**, **grupo de edad**, **causa**, la interacción **causa-sexo** y la interacción **causa-grupo de edad**. Algunos de los coeficientes de este modelo están recogidos en la [Tabla D.6](#).

Coeficientes	Estimador	Error típico	Z-value	P-valor
Intercepto	-10.75	0.27	-39.93	< 0.01
gidade2	-1.72	0.48	-3.56	< 0.01
gidade3	-1.02	0.36	-2.84	< 0.01
⋮	⋮	⋮	⋮	⋮
sexo2	-0.63	0.03	-20.66	< 0.01
as.factor(causa)II:011	-5.43	3.19	-1.70	0.09
as.factor(causa)II:012	-5.94	3.16	-1.88	0.06
⋮	⋮	⋮	⋮	⋮
gidade2:as.factor(causa)II:011	-2.71	28.07	-0.10	0.92
gidade3:as.factor(causa)II:011	-4.83	54.80	-0.09	0.93
⋮	⋮	⋮	⋮	⋮
sexo2:as.factor(causa)II:011	-0.13	0.04	-3.02	< 0.01
sexo2:as.factor(causa)II:012	-0.01	0.04	-0.25	0.81
⋮	⋮	⋮	⋮	⋮

Tabla D.6: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2020. El modelo consta de un total de 722 coeficientes.

Para la validación del modelo se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados ([Figura D.5](#)) y el segundo, es el qq-plot de los residuos de la deviance ([Figura D.6](#)). En la [Figura D.5](#) se observa que existen patrones no lineales con respecto a los valores ajustados y en la [Figura D.6](#) se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

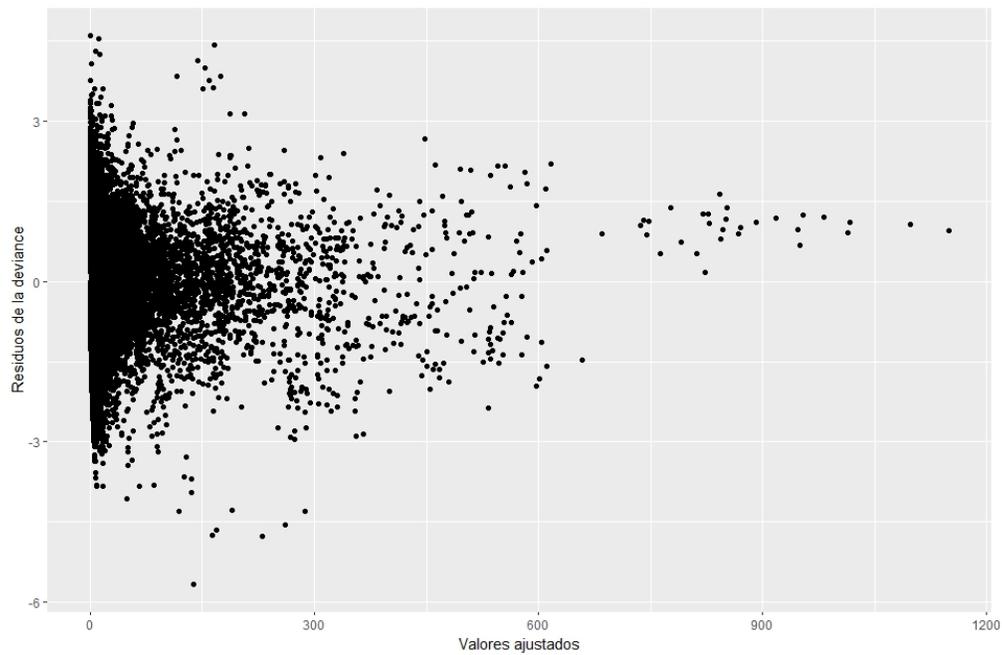


Figura D.5: Residuos de la deviance del modelo seleccionado para el año 2020 frente a los valores ajustados.

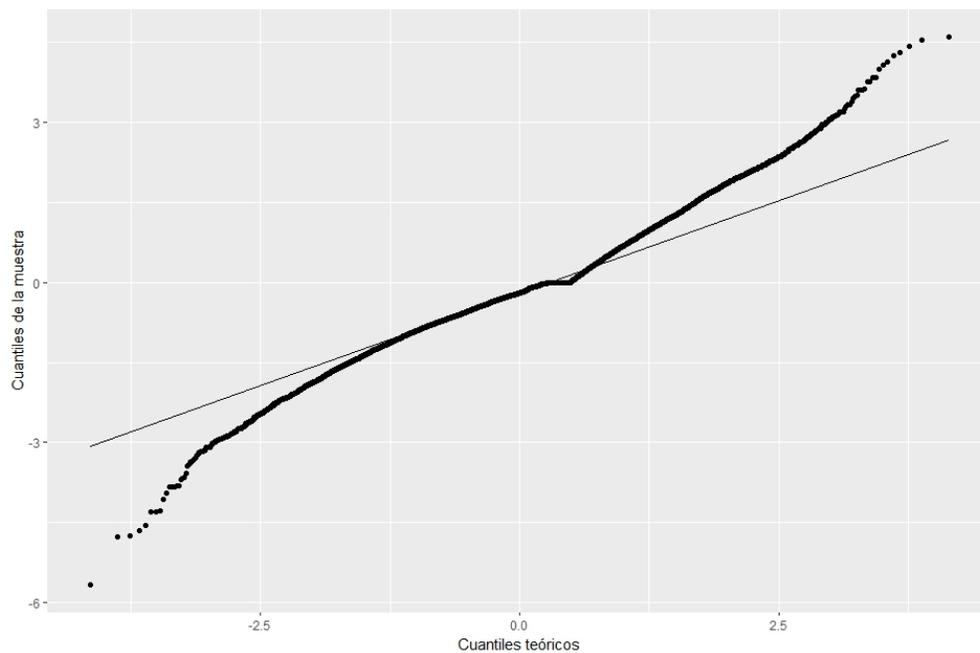


Figura D.6: QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2020.

En la [Tabla D.7](#) para cada causa seleccionada para el año 2020 están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.7](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	1988	497.3	1490.7	74.98 %
II:011	459	610.3	-151.3	-32.96 %
II:012	834	871.9	-37.9	-4.54 %
II:015	562	479.1	82.9	14.75 %
II:018	1540	1575.7	-35.7	-2.32 %
II:028	478	645	-167	-34.94 %
II:035	434	587.8	-153.8	-35.44 %
II:Resto	4176	4393.1	-217.1	-5.20 %
III	168	116.4	51.6	30.71 %
IV:044	569	706.7	-137.7	-24.20 %
IV:Resto	193	135	58	30.05 %
V:046	2351	1970.8	380.2	16.17 %
V:Resto	88	84.8	3.2	3.64 %
VI-VIII:051	795	667.9	127.1	15.99 %
VI-VIII:052	922	670	252	27.33 %
VI-VIII:Resto	4	8.8	-4.8	-120 %
IX:054	1097	740.9	356.1	32.46 %
IX:055	846	1714.8	-868.8	-102.70 %
IX:056	1256	1603.1	-347.1	-27.64 %
IX:057:I50.9	1325	1670.1	-345.1	-26.05 %
IX:057:Resto	347	523.5	-176.5	-50.86 %
IX:058	2075	2119.9	-44.9	-2.16 %
IX:059	2055	3741.8	-1686.8	-82.08 %
IX:Resto	504	845.7	-341.7	-67.80 %
X:063	496	883.3	-387.3	-78.08 %
X:064	842	1760.9	-918.9	-109.13 %
X:066:J96.0	15	23.8	-8.8	-58.67 %
X:066:J96.9	40	92.9	-52.9	-132.25 %
X:067	1659	1944.5	-285.5	-17.21 %
X:Resto	87	200.3	-113.3	-130.23 %
XI:072	929	892.3	36.7	3.95 %
XI:Resto	612	756.6	-144.6	-23.63 %
XII	149	159.1	-10.1	-6.78 %
XIII	160	252.8	-92.8	-58 %
XIV	1016	896.3	119.7	11.78 %
XV-XVII	72	87.5	-15.5	-21.53 %
XVIII	391	873.2	-482.2	-123.32 %
XX	1289	1513	-224	-17.38 %
Total	32823	37316.8	-4493.8	-13.69 %

Tabla D.7: Defunciones observadas, esperadas (a partir de los datos del período 1999-2019), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2020.

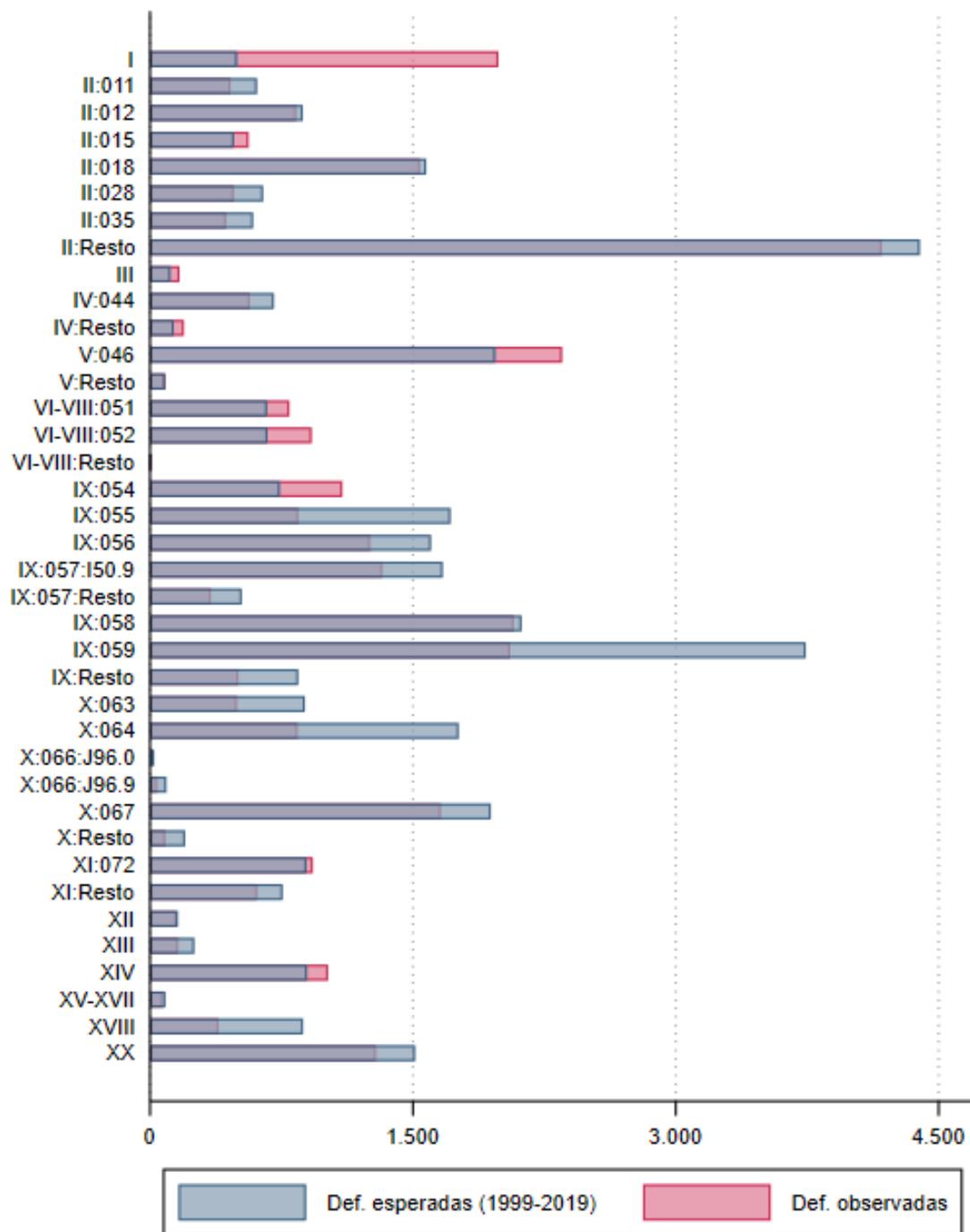


Figura D.7: Defunciones observadas y esperadas, a partir de los datos del período 1999-2019 en función de la de las causas de muerte seleccionadas. Galicia, 2020.

Agrupando por el capítulo de la causa de muerte, en la [Tabla D.8](#) están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.8](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	1988	497.3	1490.7	74.98 %
II	8483	9162.9	-679.9	-8.01 %
III	168	116.4	51.6	30.71 %
IV	762	841.7	-79.7	-10.46 %
V	2439	2055.6	383.4	15.72 %
VI-VIII	1721	1346.6	374.4	21.75 %
IX	9505	12959.8	-3454.8	-36.35 %
X	3139	4905.8	-1766.8	-56.29 %
XI	1541	1648.9	-107.9	-7 %
XII	149	159.1	-10.1	-6.78 %
XIII	160	252.8	-92.8	-58 %
XIV	1016	896.3	119.7	11.78 %
XV-XVII	72	87.5	-15.5	-21.53 %
XVIII	391	873.2	-482.2	-123.32 %
XX	1289	1513	-224	-17.38 %
Total	32823	37316.8	-4493.8	-13.69 %

Tabla D.8: Defunciones observadas, esperadas (a partir de los datos del período 1999-2019), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2020.

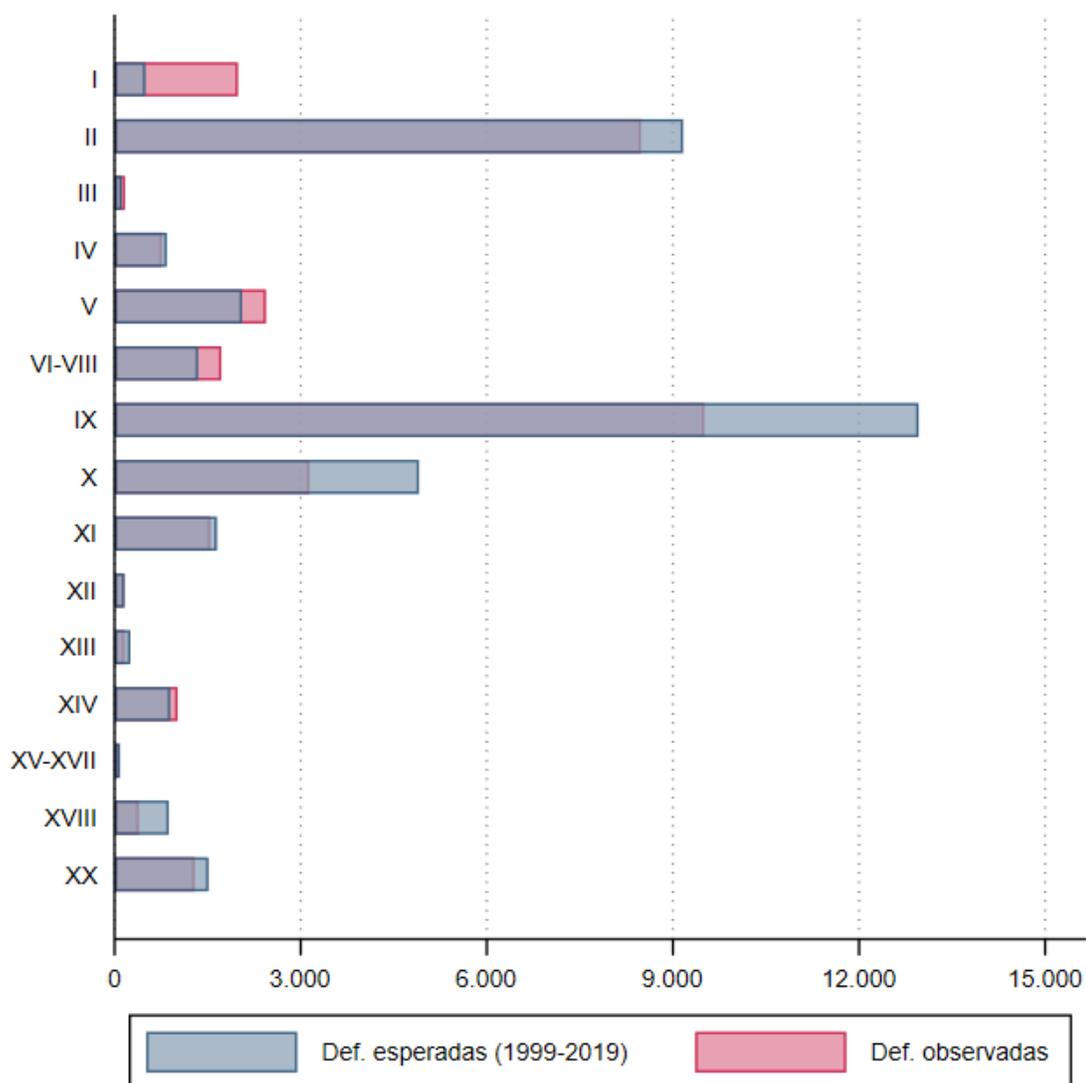


Figura D.8: Defunciones observadas y esperadas, a partir de los datos del período 1999-2019 en función del capítulo de la causa de muerte. Galicia, 2020.

Ahora, para el ajuste del modelo del año 2021 se siguieron los mismos pasos que en los dos años anteriores. En este caso, se tomó como período histórico de referencia el 1999-2020. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. El valor del AIC para el modelo sin efectos aleatorios fue 186421.7 y el del modelo con efectos aleatorios fue 181346.3 (Tabla D.9). Por lo tanto, el modelo con menor AIC es el modelo con efectos aleatorios.

Variables explicativas del modelo	AIC
<ul style="list-style-type: none"> ■ Población expuesta. 	186421.7
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variable grupo: causa. 	181346.3
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	138733.5
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad y causa. ■ Variable grupo: causa. ■ Componente ar1. 	138669.4
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-sexo. ■ Variable grupo: causa. ■ Componente ar1. 	136149.6
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	117939.9
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa, interacción causa-sexo e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	109211
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Pendiente aleatoria: sexo. ■ Componente ar1. 	136402.9

Tabla D.9: Valores del AIC de los distintos modelos mixtos ajustados para el año 2021. Estos resultados se obtienen con la función AIC de .

A partir del modelo con efectos aleatorios se ajustaron modelos con intercepto aleatorio tomando como variables explicativas las mismas que para los dos años anteriores. También se ajustaron modelos con intercepto y pendiente aleatorios. De estos modelos, como ocurrió con el 2019 y el 2020, el único que convergió fue el modelo con:

- Variables explicativas: **sexo** y **grupo de edad**.
- Población expuesta
- Variable grupo: **causa**.
- Pendiente aleatoria: **sexo**.

Los demás modelos con intercepto y pendiente aleatorios no convergieron, ya que todos presentaron el mismo error relacionado con que la matriz hessiana no era definida positiva. Como se indica en la [Tabla D.9](#), en algunos modelos se ha considerado una componente autorregresiva, **ar1**, para incluir la dependencia temporal.

Por tanto, en la [Tabla D.9](#) se puede observar que el menor valor del AIC es 109211 que se corresponde con el modelo que tiene como variables explicativas: **sexo**, **grupo de edad**, **causa**, la interacción **causa-sexo** y la interacción **causa-grupo de edad**. Algunos de los coeficientes de este modelo están recogidos en la [Tabla D.10](#).

Coeficientes	Estimador	Error típico	Z-value	P-valor
Intercepto	-10.86	0.28	-39.13	< 0.01
gidade2	-1.7	0.49	-3.54	< 0.01
gidade3	-0.96	0.35	-2.70	< 0.01
⋮	⋮	⋮	⋮	⋮
sexo2	-0.65	0.03	-21.83	< 0.01
as.factor(causa)II:011	-5.16	2.89	-1.78	0.07
as.factor(causa)II:012	-5.67	2.87	-1.98	0.05
⋮	⋮	⋮	⋮	⋮
gidade2:as.factor(causa)II:011	-2.49	23.07	-0.11	0.91
gidade3:as.factor(causa)II:011	-4.52	41.27	-0.11	0.91
⋮	⋮	⋮	⋮	⋮
sexo2:as.factor(causa)II:011	-0.11	0.04	-2.56	0.01
sexo2:as.factor(causa)II:012	0.02	0.04	0.41	0.68
⋮	⋮	⋮	⋮	⋮

Tabla D.10: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2021. El modelo consta de un total de 722 coeficientes.

Para la validación del modelo se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados ([Figura D.9](#)) y el segundo, es el qq-plot de los residuos de la deviance ([Figura D.10](#)). En la [Figura D.9](#) se observa que existen patrones no lineales con respecto a los valores ajustados y en la [Figura D.10](#) se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

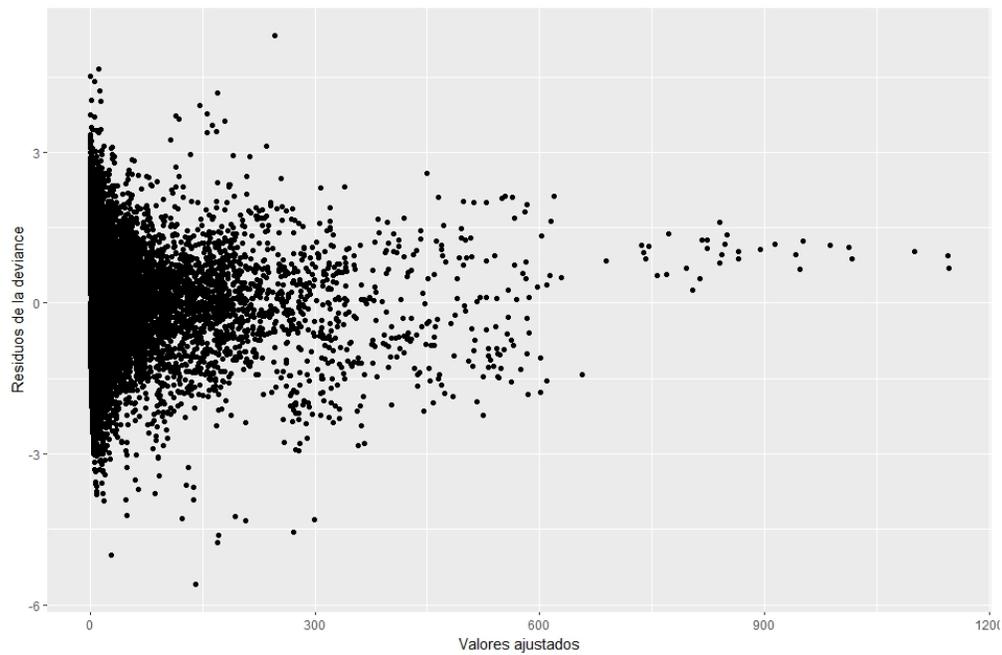


Figura D.9: Residuos de la deviance del modelo seleccionado para el año 2021 frente a los valores ajustados.

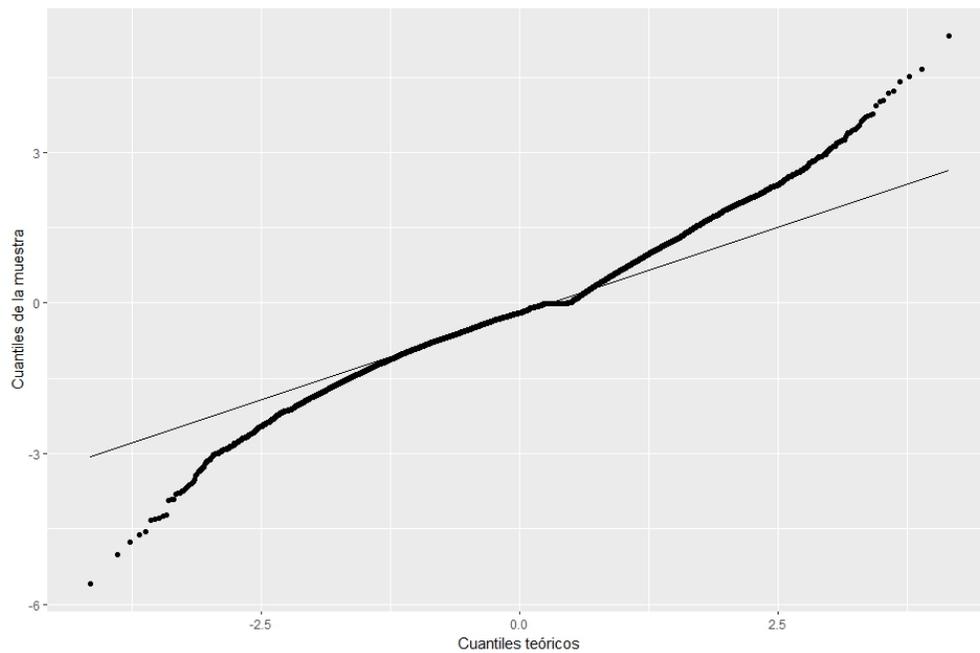


Figura D.10: QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2021.

En la [Tabla D.11](#) para cada causa seleccionada para el año 2020 están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.11](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	1954	939.4	1014.6	51.92 %
II:011	379	625.3	-246.3	-64.99 %
II:012	822	874.4	-52.4	-6.37 %
II:015	521	499.9	21.1	4.05 %
II:018	1586	1618.2	-32.2	-2.03 %
II:028	496	640.8	-144.8	-29.19 %
II:035	444	598.5	-154.5	-34.80 %
II:Resto	4216	4443.4	-227.4	-5.39 %
III	168	120.1	47.9	28.51 %
IV:044	618	754.3	-136.3	-22.06 %
IV:Resto	282	146.5	135.5	48.05 %
V:046	2154	1965.0	189	8.77 %
V:Resto	73	88.9	-15.9	-21.78 %
VI-VIII:051	667	660.1	6.9	1.03 %
VI-VIII:052	897	681.4	215.6	24.04 %
VI-VIII:Resto	9	8	1	11.11 %
IX:054	1120	890.2	229.8	20.52 %
IX:055	851	1757.8	-906.8	-106.56 %
IX:056	1188	1678.8	-490.8	-41.31 %
IX:057:I50.9	1441	1696.6	-255.6	-17.74 %
IX:057:Resto	360	594.1	-234.1	-65.03 %
IX:058	2075	2087.6	-12.6	-0.61 %
IX:059	2056	3625.4	-1569.4	-76.33 %
IX:Resto	490	847.3	-357.3	-72.92 %
X:063	443	865.4	-422.4	-95.35 %
X:064	749	1752.4	-1003.4	-133.97 %
X:066:J96.0	24	28.4	-4.4	-18.33 %
X:066:J96.9	70	103.6	-33.6	-48 %
X:067	1505	1962.2	-457.2	-30.38 %
X:Resto	94	181.9	-87.9	-93.51 %
XI:072	974	916.2	57.8	5.93 %
XI:Resto	615	767.7	-152.7	-24.83 %
XII	148	163.8	-15.8	-10.68 %
XIII	220	245.2	-25.2	-11.45 %
XIV	1128	925.8	202.2	17.93 %
XV-XVII	88	84.8	3.2	3.64 %
XVIII	488	623.9	-135.9	-27.85 %
XX	1440	1570.1	-130.1	-9.03 %
Total	32853	38033.2	-5180.2	-15.77 %

Tabla D.11: Defunciones observadas, esperadas (a partir de los datos del período 1999-2020), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2021.

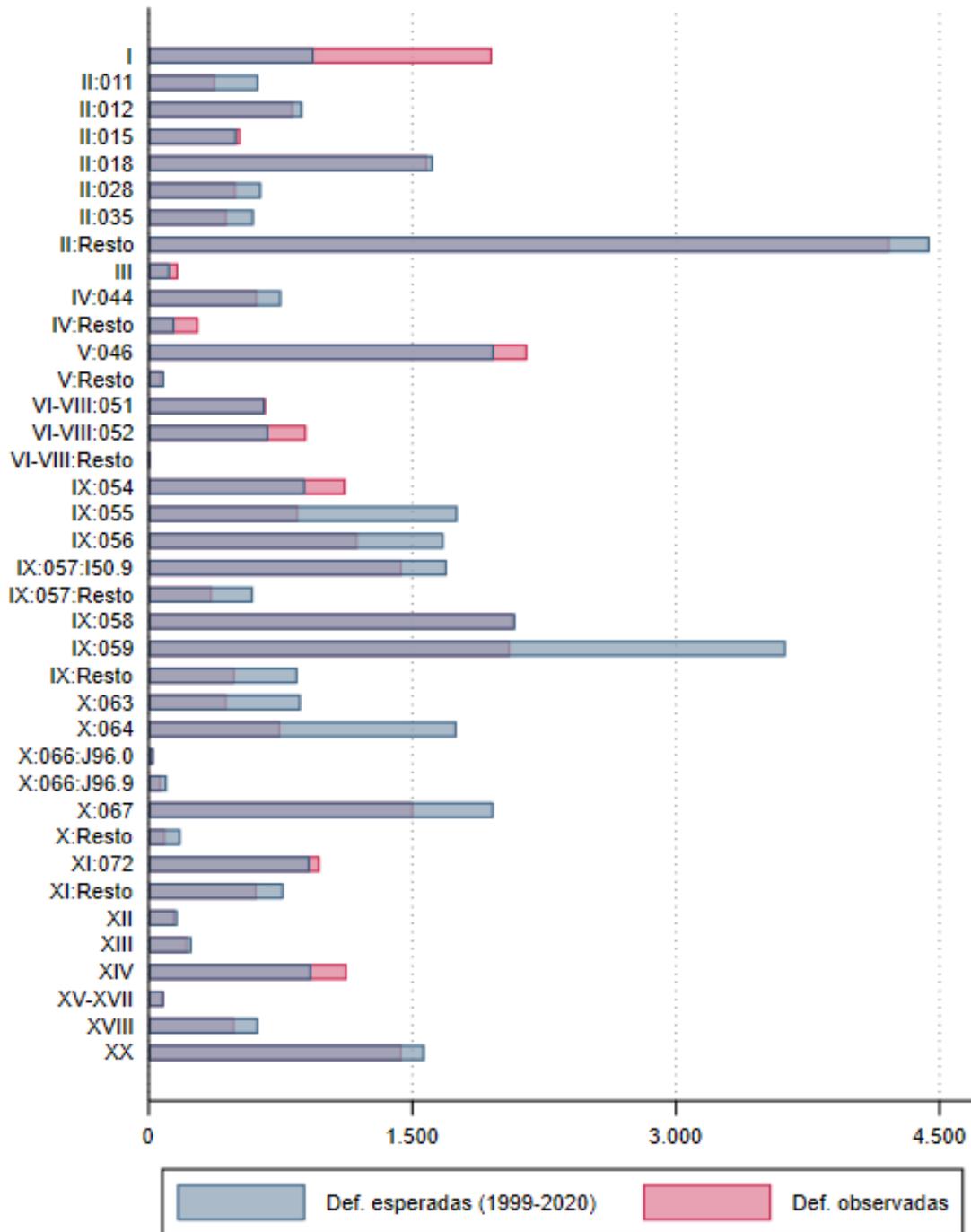


Figura D.11: Defunciones observadas y esperadas, a partir de los datos del período 1999-2020 en función de la de las causas de muerte seleccionadas. Galicia, 2021.

Agrupando por el capítulo de la causa de muerte, en la [Tabla D.12](#) están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.12](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	1954	939.4	1014.6	51.92 %
II	8464	9300.5	-836.5	-9.88 %
III	168	120.1	47.9	28.51 %
IV	900	900.7	-0.7	-0.08 %
V	2227	2053.8	173.2	7.78 %
VI-VIII	1573	1349.4	223.6	14.21 %
IX	9581	13177.9	-3596.9	-37.54 %
X	2885	4894	-2009	-69.64 %
XI	1589	1683.9	-94.9	-5.97 %
XII	148	163.8	-15.8	-10.68 %
XIII	220	245.2	-25.2	-11.45 %
XIV	1128	925.8	202.2	17.93 %
XV-XVII	88	84.8	3.2	3.64 %
XVIII	488	623.9	-135.9	-27.85 %
XX	1440	1570.1	-130.1	-9.03 %
Total	32853	38033.2	-5180.2	-15.77 %

Tabla D.12: Defunciones observadas, esperadas (a partir de los datos del período 1999-2020), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2021.

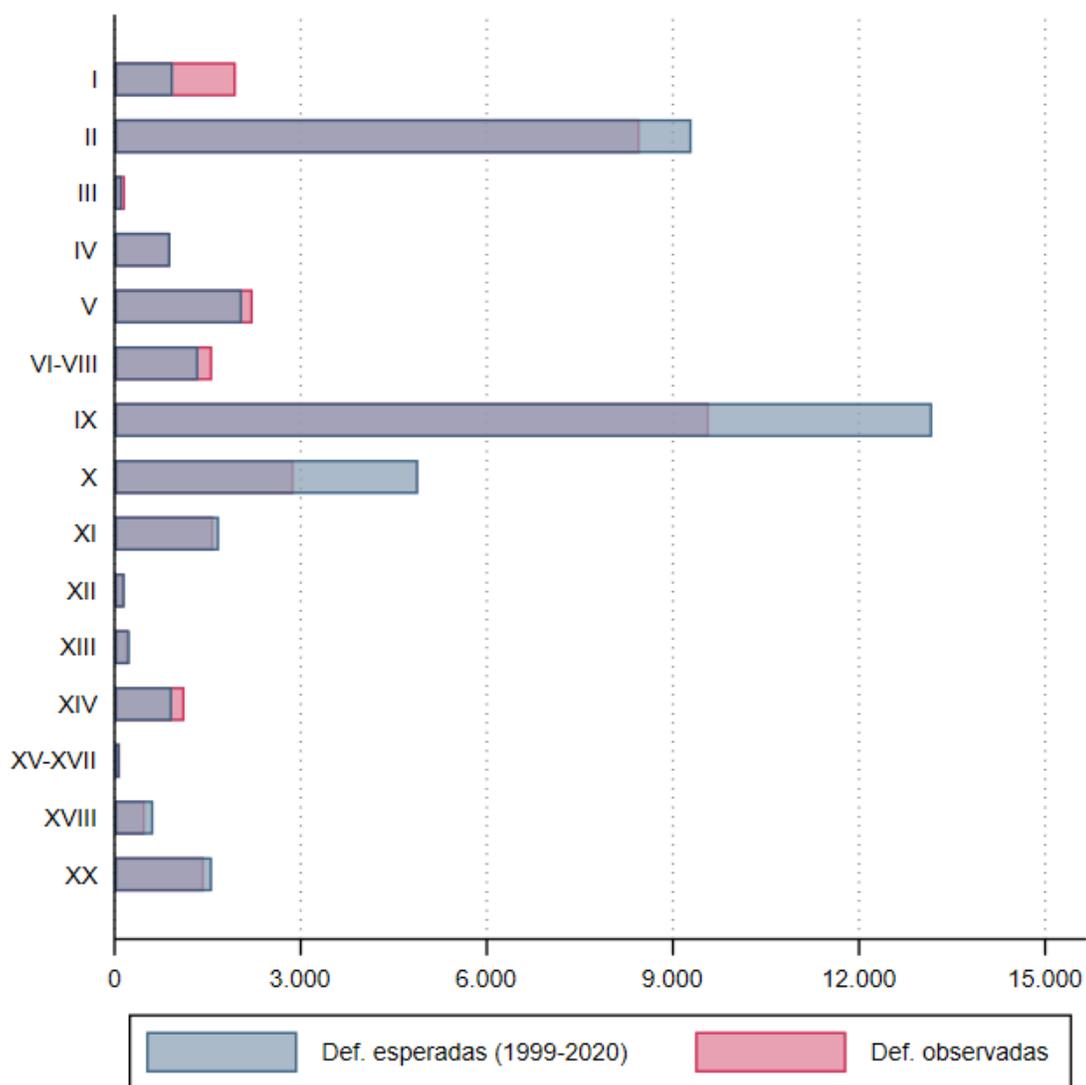


Figura D.12: Defunciones observadas y esperadas, a partir de los datos del período 1999-2020 en función del capítulo de la causa de muerte. Galicia, 2021.

Por último, para el ajuste del modelo del año 2022 se siguió el mismo procedimiento que en los años anteriores. En este caso, se tomó como período histórico de referencia el 1999-2021. En primer lugar, se comprobó si era necesario añadir efectos aleatorios. El valor del AIC para el modelo sin efectos aleatorios fue 194702.5 y el del modelo con efectos aleatorios fue 189419.8 (Tabla D.13). Por lo tanto, el modelo con menor AIC es el modelo con efectos aleatorios.

Variables explicativas del modelo	AIC
<ul style="list-style-type: none"> ■ Población expuesta. 	194702.5
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variable grupo: causa. 	189419.8
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	144663.2
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad y causa. ■ Variable grupo: causa. ■ Componente ar1. 	144598.9
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-sexo. ■ Variable grupo: causa. ■ Componente ar1. 	141973.6
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	123153.5
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo, grupo de edad, causa, interacción causa-sexo e interacción causa-grupo de edad. ■ Variable grupo: causa. ■ Componente ar1. 	114192.7
<ul style="list-style-type: none"> ■ Población expuesta. ■ Variables explicativas: sexo y grupo de edad. ■ Variable grupo: causa. ■ Pendiente aleatoria: sexo. ■ Componente ar1. 	142229.4

Tabla D.13: Valores del AIC de los distintos modelos mixtos ajustados para el año 2022. Estos resultados se obtienen con la función AIC de .

Ahora, a partir del modelo con efectos aleatorios se ajustaron modelos con intercepto aleatorio tomando como variables explicativas las mismas que para los años anteriores. También se ajustaron modelos con intercepto y pendiente aleatorios. De estos modelos, como ocurrió con los años anteriores, el único que convergió fue el modelo con:

- Variables explicativas: **sexo** y **grupo de edad**.
- Población expuesta
- Variable grupo: **causa**.
- Pendiente aleatoria: **sexo**.

Los demás modelos con intercepto y pendiente aleatorios no convergieron, ya que todos presentaron el mismo error relacionado con que la matriz hessiana no era definida positiva. Como se indica en la [Tabla D.13](#), en algunos modelos se ha considerado una componente autorregresiva, **ar1**, para incluir la dependencia temporal.

Por tanto, en la [Tabla D.13](#) se puede observar que el menor valor del AIC es 114192.7 que se corresponde con el modelo que tiene como variables explicativas: **sexo**, **grupo de edad**, **causa**, la interacción **causa-sexo** y la interacción **causa-grupo de edad**. Algunos de los coeficientes de este modelo están recogidos en la [Tabla D.14](#).

Coeficientes	Estimador	Error típico	Z-value	P-valor
Intercepto	-10.85	0.28	-39.36	< 0.01
gidade2	-1.78	0.48	-3.67	< 0.01
gidade3	-1.11	0.36	-3.05	< 0.01
⋮	⋮	⋮	⋮	⋮
sexo2	-0.67	0.03	-22.89	< 0.01
as.factor(causa)II:011	-5.56	3.47	-1.60	0.11
as.factor(causa)II:012	-6.09	3.46	-1.76	0.08
⋮	⋮	⋮	⋮	⋮
gidade2:as.factor(causa)II:011	-2.67	30.74	-0.09	0.93
gidade3:as.factor(causa)II:011	-4.80	61.29	-0.08	0.94
⋮	⋮	⋮	⋮	⋮
sexo2:as.factor(causa)II:011	-0.10	0.04	-2.28	0.02
sexo2:as.factor(causa)II:012	0.03	0.04	0.65	0.51
⋮	⋮	⋮	⋮	⋮

Tabla D.14: Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2022. El modelo consta de un total de 722 coeficientes.

Para la validación del modelo se elaboraron dos gráficos: el primero representa los residuos de la deviance frente a los valores ajustados ([Figura D.13](#)) y el segundo, es el qq-plot de los residuos de la deviance ([Figura D.14](#)). En la [Figura D.13](#) se observa que existen patrones no lineales con respecto a los valores ajustados y en la [Figura D.14](#) se identifican puntos alejados de la diagonal por lo tanto los residuos de la deviance no parecen seguir una distribución normal.

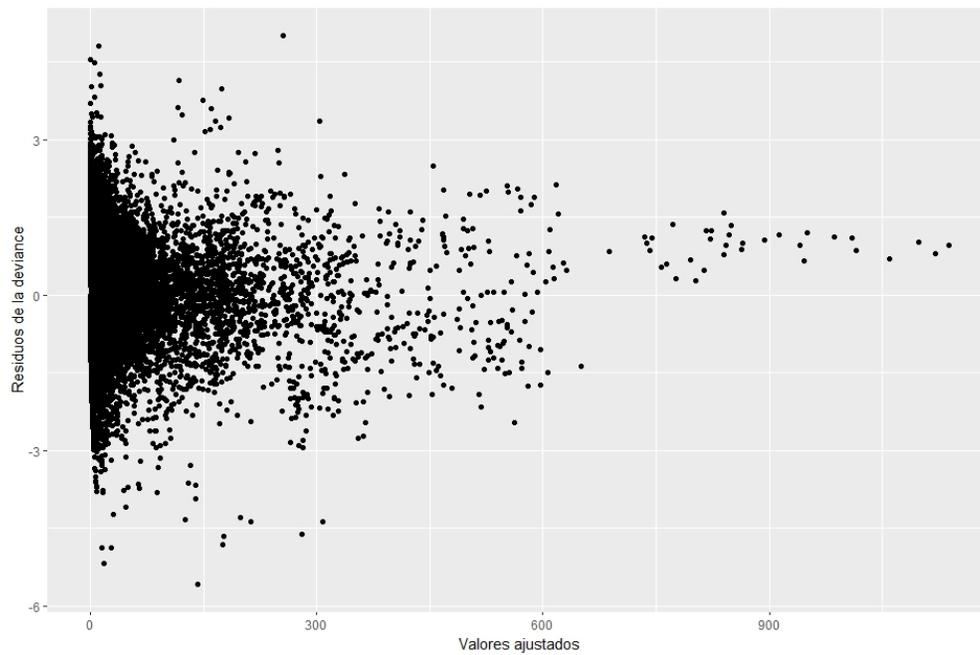


Figura D.13: Residuos de la deviance del modelo seleccionado para el año 2022 frente a los valores ajustados.

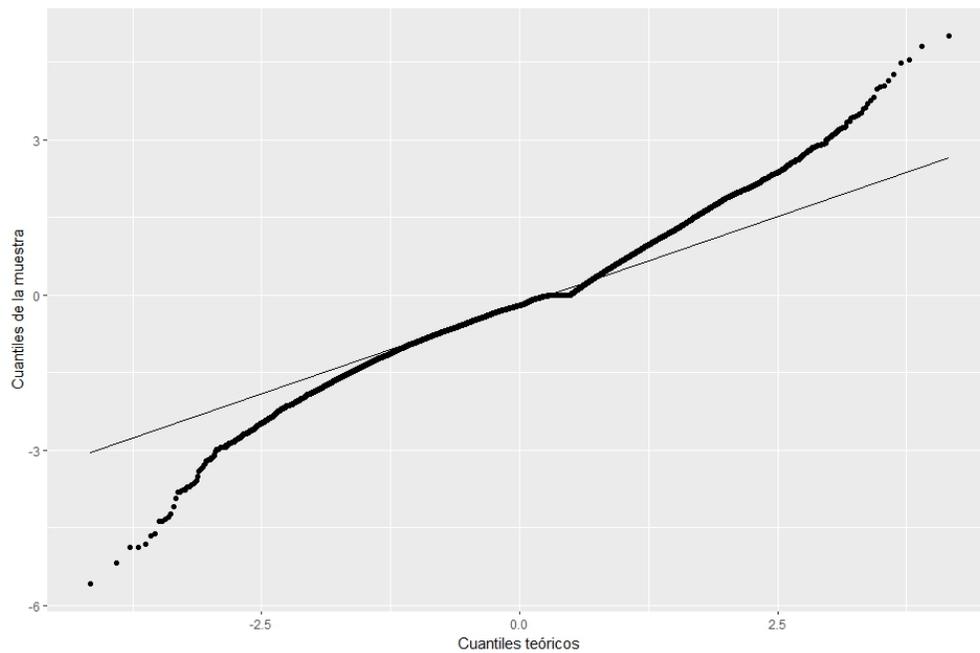


Figura D.14: QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2022.

En la [Tabla D.15](#) para cada causa seleccionada para el año 2020 están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.15](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por causa	Porcentaje de exceso
I	2291	1069.1	1221.9	53.33 %
II:011	381	589.3	-208.3	-54.67 %
II:012	793	879.9	-86.9	-10.96 %
II:015	533	493.8	39.2	7.35 %
II:018	1547	1650.8	-103.8	-6.71 %
II:028	462	662.9	-200.9	-43.48 %
II:035	409	603.8	-194.8	-47.63 %
II:Resto	4199	4423.7	-224.7	-5.35 %
III	159	124.6	34.4	21.64 %
IV:044	808	808.5	-0.5	-0.06 %
IV:Resto	349	182.1	166.9	47.82 %
V:046	2401	1897.8	503.2	20.96 %
V:Resto	79	84.5	-5.5	-6.96 %
VI-VIII:051	675	657.3	17.7	2.62 %
VI-VIII:052	950	684.2	265.8	27.98 %
VI-VIII:Resto	13	8.5	4.5	34.62 %
IX:054	1109	956.1	152.9	13.79 %
IX:055	751	1778.1	-1027.1	-136.76 %
IX:056	1125	1742.2	-617.2	-54.86 %
IX:057:I50.9	1468	1756.6	-288.6	-19.66 %
IX:057:Resto	318	625.9	-307.9	-96.82 %
IX:058	1917	2106.3	-189.3	-9.87 %
IX:059	1935	3642.2	-1707.2	-88.23 %
IX:Resto	536	833.7	-297.7	-55.54 %
X:063	619	816.1	-197.1	-31.84 %
X:064	774	1632.5	-858.5	-110.92 %
X:066:J96.0	27	35.8	-8.8	-32.59 %
X:066:J96.9	121	126.4	-5.4	-4.46 %
X:067	1885	1910.4	-25.4	-1.35 %
X:Resto	82	177.3	-95.3	-116.22 %
XI:072	1002	930.7	71.3	7.12 %
XI:Resto	583	779.9	-196.9	-33.77 %
XII	159	166.8	-7.8	-4.91 %
XIII	280	264.8	15.2	5.43 %
XIV	1236	977.5	258.5	20.91 %
XV-XVII	63	91.8	-28.8	-45.71 %
XVIII	1034	633.9	400.1	38.69 %
XX	1414	1649.6	-235.6	-16.66 %
Total	34487	38455.2	-3968.2	-11.51 %

Tabla D.15: Defunciones observadas, esperadas (a partir de los datos del período 1999-2021), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2022.

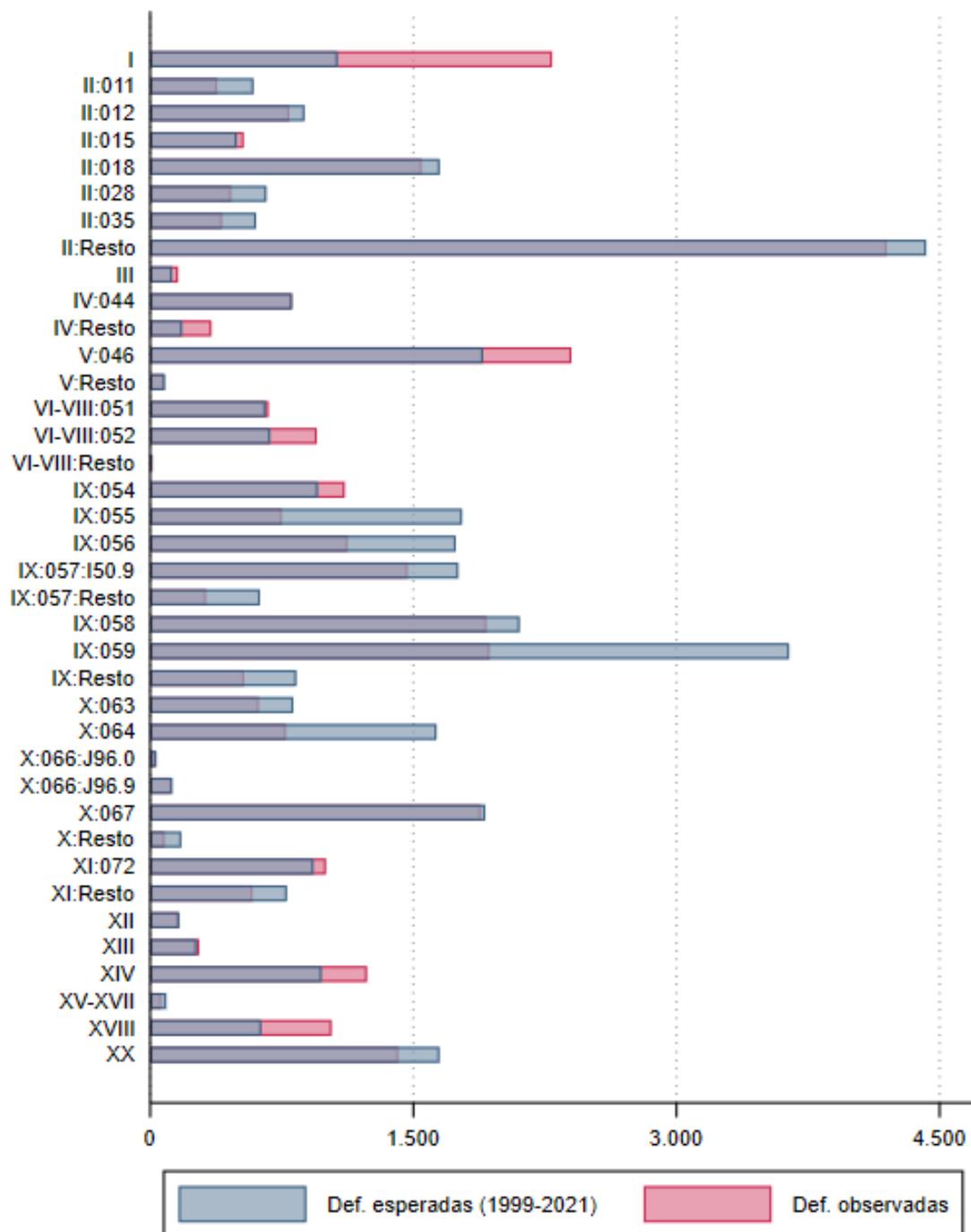


Figura D.15: Defunciones observadas y esperadas, a partir de los datos del período 1999-2021 en función de la de las causas de muerte seleccionadas. Galicia, 2022.

Agrupando por el capítulo de la causa de muerte, en la [Tabla D.16](#) están recogidas las defunciones observadas, las esperadas, el exceso de mortalidad y el porcentaje de exceso sobre el total de defunciones observadas. Estos valores se pueden ver gráficamente en la [Figura D.16](#).

Causa de defunción	Defunciones observadas	Defunciones estimadas	Exceso por capítulo	Porcentaje de exceso
I	2291	1069.1	1221.9	53.33 %
II	8324	9304.1	-980.1	-11.77 %
III	159	124.6	34.4	21.64 %
IV	1157	990.6	166.4	14.38 %
V	2480	1982.3	497.7	20.07 %
VI-VIII	1638	1350	288	17.58 %
IX	9159	13441	-4282	-46.75 %
X	3508	4698.4	-1190.4	-33.93 %
XI	1585	1710.6	-125.6	-7.92 %
XII	159	166.8	-7.8	-4.91 %
XIII	280	264.8	15.2	5.43 %
XIV	1236	977.5	258.5	20.91 %
XV-XVII	63	91.8	-28.8	-45.71 %
XVIII	1034	633.9	400.1	38.69 %
XX	1414	1649.6	-235.6	-16.66 %
Total	34487	38455.2	-3968.2	-11.51 %

Tabla D.16: Defunciones observadas, esperadas (a partir de los datos del período 1999-2021), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2022.

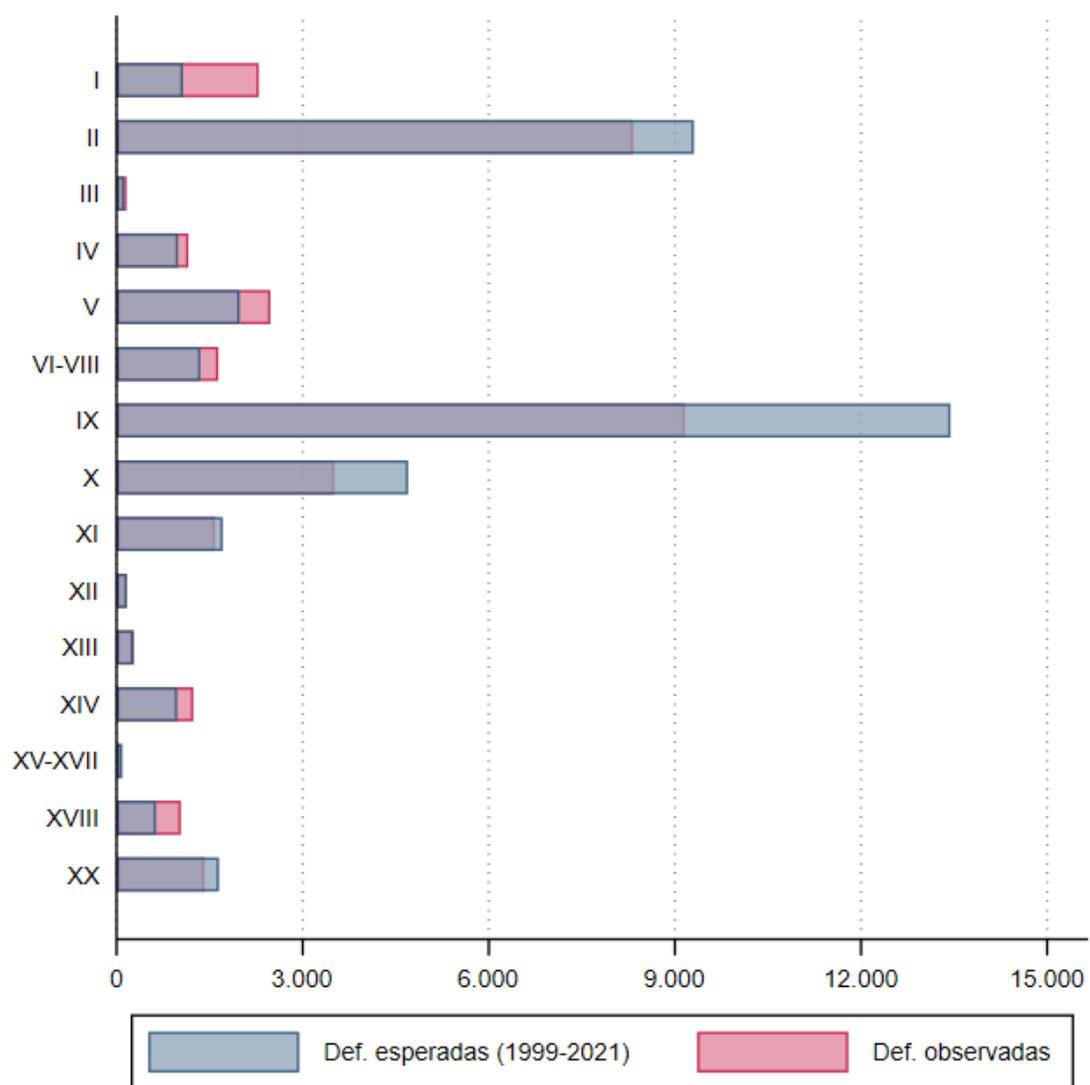


Figura D.16: Defunciones observadas y esperadas, a partir de los datos del período 1999-2021 en función del capítulo de la causa de muerte. Galicia, 2022.

Bibliografía

- [1] Achebak, H., Devolder, D., y Ballester, J. (2019). Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: A national time-series analysis. *The Lancet Planetary Health*, 3(7):e297–e306.
- [2] Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley, 2 edition.
- [3] Blangiardo, M., Cameletti, M., Pirani, M., Corsetti, G., Battaglini, M., y Baio, G. (2020). Estimating weekly excess mortality at sub-national level in Italy during the COVID-19 pandemic. *PLOS ONE*, 15(10):e0240286.
- [4] Centro Nacional de Epidemiología del Instituto de Salud Carlos III (CNE). Sistema de monitorización de la mortalidad (MoMo). https://momo.isciii.es/panel_momo/#section-momo. Accedido 3 de octubre de 2023.
- [5] European mortality monitoring (EuroMOMO). <https://euromomo.eu/>. Accedido 19 de octubre de 2023.
- [6] Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC, 2 edition.
- [7] Foreman, K. J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., McGaughey, M., et al. (2018). Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: Reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392(10159):2052–2090.
- [8] Islam, N., Shkolnikov, V. M., Acosta, R. J., Klimkin, I., Kawachi, I., Irizarry, R. A., et al. (2021). Excess deaths associated with covid-19 pandemic in 2020: Age and sex disaggregated time series analysis in 29 high income countries. *BMJ (Clinical research ed.)*, 373(n1137).
- [9] Organización Mundial de la Salud (OMS) (1993). La CIE-10 clasificación de trastornos mentales y del comportamiento. *Organización Mundial de la Salud (OMS)*.
- [10] Santos, A. M. D., Souza, B. F. D., Carvalho, C. A. D., Campos, M. A. G., Oliveira, B. L. C. A. D., Diniz, E. M., et al. (2021). Excess deaths from all causes and by COVID-19 in Brazil in 2020. *Revista de Saúde Pública*, 55:71.
- [11] Shin, M. S., Sim, B., Jang, W. M., y Lee, J. Y. (2021). Estimation of excess all-cause mortality during COVID-19 pandemic in Korea. *Journal of Korean Medical Science*, 36(39):e280.
- [12] Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Taylor & Francis Group, Bosa Roca, UNITED STATES.
- [13] Tawiah, K., Iddrisu, W. A., y Asampana Asosega, K. (2021). Zero-Inflated time series modelling of COVID-19 deaths in Ghana. *Journal of Environmental and Public Health*, 2021:1–9.
- [14] Vélez Ibarrola, R. y García Pérez, A. (2012). *Principios de Inferencia Estadística*. UNED - Universidad Nacional de Educación a Distancia.
- [15] Verbeeck, J., Faes, C., Neyens, T., Hens, N., Verbeke, G., Deboosere, P., et al. (2023). A linear mixed model to estimate COVID-19-induced excess mortality. *Biometrics*, 79(1):417–425.

- [16] Vieira, A., Peixoto, V. R., Aguiar, P., y Abrantes, A. (2020). Rapid estimation of excess mortality during the COVID-19 pandemic in Portugal. *Journal of Epidemiology and Global Health*, 10(3):209.
- [17] Villaseñor Alva, J. A., Hernández de la Rosa, F. A., y Vaquera Huerta, H. (2001). La prueba de la razón de verosimilitudes para la distribución Poisson contra binomial negativa o binomial. *Agrociencia*, 35(2):211–222.
- [18] Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., y Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer.

Índice de Tablas

1.1. Defunciones anuales observadas, estimadas, exceso por todas las causas y porcentaje de excesos anuales sobre el total de defunciones anuales observadas. Galicia, 2015-2022. Fuente: MoMo. . .	1
3.1. Definición detallada y código de la CIE-10 para cada grupo de causas de defunción seleccionado.	24
3.2. Media y varianza de la variable defunciones para cada una de las causas seleccionadas, a partir de los datos estructurados para el ajuste del modelo. Galicia, 1999-2018.	26
3.3. Parámetro de sobre-dispersión, estadístico χ^2 del contraste de sobre-dispersión y su p-valor asociado para cada causa seleccionada y tomando la variable año como continua y como categórica. Estos resultados se obtienen al ajustar el modelo de regresión Binomial Negativa en Stata. El símbolo * representa que el modelo no convergió para esa causa.	28
3.4. Descripción de los cuatro modelos mixtos ajustados.	29
3.5. Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2019. Estos resultados se obtienen con la función AIC de \mathbb{R} . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.	30
3.6. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2019.	31
3.7. Defunciones observadas, esperadas (a partir de los datos del período 2009-2018), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2019.	33
3.8. Defunciones observadas, esperadas (a partir de los datos del período 2009-2018), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2019.	35
3.9. Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2020. Estos resultados se obtienen con la función AIC de \mathbb{R} . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.	37
3.10. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2020.	38
3.11. Defunciones observadas, esperadas (a partir de los datos del período 2010-2019), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2020.	40
3.12. Defunciones observadas, esperadas (a partir de los datos del período 2010-2019), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2020.	42
3.13. Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2021. Estos resultados se obtienen con la función AIC de \mathbb{R} . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.	44
3.14. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2021.	45
3.15. Defunciones observadas, esperadas (a partir de los datos del período 2011-2020), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2021.	47
3.16. Defunciones observadas, esperadas (a partir de los datos del período 2011-2020), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2021.	49

3.17. Valores del AIC para las causas de defunción de los distintos modelos mixtos ajustados para el año 2022. Estos resultados se obtienen con la función AIC de R . En negrita se resaltan los menores valores del AIC para cada una de las causas de defunción.	51
3.18. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes de los modelos ajustados para el año 2022.	52
3.19. Defunciones observadas, esperadas (a partir de los datos del período 2012-2021), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2022.	54
3.20. Defunciones observadas, esperadas (a partir de los datos del período 2012-2021), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2022.	56
D.1. Valores del AIC de los distintos modelos mixtos ajustados para el año 2019. Estos resultados se obtienen con la función AIC de R	72
D.2. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2019. El modelo consta de un total de 722 coeficientes.	73
D.3. Defunciones observadas, esperadas (a partir de los datos del período 1999-2018), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2019.	75
D.4. Defunciones observadas, esperadas (a partir de los datos del período 1999-2018), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2019.	77
D.5. Valores del AIC de los distintos modelos mixtos ajustados para el año 2020. Estos resultados se obtienen con la función AIC de R	79
D.6. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2020. El modelo consta de un total de 722 coeficientes.	80
D.7. Defunciones observadas, esperadas (a partir de los datos del período 1999-2019), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2020.	82
D.8. Defunciones observadas, esperadas (a partir de los datos del período 1999-2019), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2020.	84
D.9. Valores del AIC de los distintos modelos mixtos ajustados para el año 2021. Estos resultados se obtienen con la función AIC de R	86
D.10. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2021. El modelo consta de un total de 722 coeficientes.	87
D.11. Defunciones observadas, esperadas (a partir de los datos del período 1999-2020), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2021.	89
D.12. Defunciones observadas, esperadas (a partir de los datos del período 1999-2020), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2021.	91
D.13. Valores del AIC de los distintos modelos mixtos ajustados para el año 2022. Estos resultados se obtienen con la función AIC de R	93
D.14. Salida parcial del estimador, error típico, z-value y su p-valor de los coeficientes del modelo ajustado para el año 2022. El modelo consta de un total de 722 coeficientes.	94
D.15. Defunciones observadas, esperadas (a partir de los datos del período 1999-2021), exceso de mortalidad por causa y porcentaje de excesos sobre el total de defunciones anuales observadas en función de las causas de muerte seleccionadas. Galicia, 2022.	96
D.16. Defunciones observadas, esperadas (a partir de los datos del período 1999-2021), exceso de mortalidad por capítulo y porcentaje de excesos sobre el total de defunciones anuales observadas en función del capítulo de las causas de muerte seleccionadas. Galicia, 2022.	98

Índice de Figuras

1.1. Evolución de la tasa bruta de mortalidad y la tasa ajustada por edad (población estándar: censo del 2011) en Galicia en el período 1980-2022.	2
1.2. Evolución de la mortalidad observada en Galicia desde el año 2015 al 2022 y de la mortalidad esperada desde el año 2020 al 2022. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	3
1.3. Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función del capítulo de la causa de muerte.	4
1.4. Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función de la lista reducida de causas de muerte del capítulo <i>I</i>	5
1.5. Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 en función de la lista reducida de causas de muerte del capítulo <i>XVIII</i>	6
1.6. Defunciones observadas y esperadas, a partir de los datos del período 2015-2019, en Galicia en 2020, 2021 y 2022 para el conjunto de las causas mal definidas.	7
3.1. Evolución de la mortalidad observada en Galicia por COVID-19 desde el año 2020 al 2022. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	25
3.2. Izquierda: Residuos de la deviance del modelo seleccionado para la causa <i>I</i> frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa <i>I</i> . Año 2019.	32
3.3. Evolución de la mortalidad esperada (a partir de los datos del período 2009-2018) y observada en Galicia desde el año 2009 al 2019. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	32
3.4. Defunciones observadas y esperadas, a partir de los datos del período 2009-2018 en función del capítulo. Galicia, 2019.	34
3.5. Defunciones observadas y esperadas, a partir de los datos del período 2009-2018 en función de las causas de muerte seleccionadas. Galicia, 2019.	36
3.6. Izquierda: Residuos de la deviance del modelo seleccionado para la causa <i>I</i> frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa <i>I</i> . Año 2020.	39
3.7. Evolución de la mortalidad esperada (a partir de los datos del período 2010-2019) y observada en Galicia desde el año 2010 al 2020. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	39
3.8. Defunciones observadas y esperadas, a partir de los datos del período 2010-2019 en función del capítulo. Galicia, 2020.	41
3.9. Defunciones observadas y esperadas, a partir de los datos del período 2010-2019 en función de la de las causas de muerte seleccionadas. Galicia, 2020.	43
3.10. Izquierda: Residuos de la deviance del modelo seleccionado para la causa <i>I</i> frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa <i>I</i> . Año 2021.	46
3.11. Evolución de la mortalidad esperada (a partir de los datos del período 2011-2020) y observada en Galicia desde el año 2011 al 2021. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	46

3.12. Defunciones observadas y esperadas, a partir de los datos del período 2011-2020 en función del capítulo. Galicia, 2021.	48
3.13. Defunciones observadas y esperadas, a partir de los datos del período 2011-2020 en función de las causas de muerte seleccionadas. Galicia, 2021.	50
3.14. Izquierda: Residuos de la deviance del modelo seleccionado para la causa <i>I</i> frente a los valores ajustados. Derecha: QQ-plot de los residuos de la deviance del modelo seleccionado para la causa <i>I</i> . Año 2022.	53
3.15. Evolución de la mortalidad esperada (a partir de los datos del período 2012-2021) y observada en Galicia desde el año 2012 al 2022. En el eje <i>X</i> están representados los años y en el eje <i>Y</i> el número de defunciones.	53
3.16. Defunciones observadas y esperadas, a partir de los datos del período 2012-2021 en función del capítulo. Galicia, 2022.	55
3.17. Defunciones observadas y esperadas, a partir de los datos del período 2012-2021 en función de la de las causas de muerte seleccionadas. Galicia, 2022.	57
A.1. Captura de pantalla del Panel MoMo con los datos de Galicia en el período 2015-2022.	62
A.2. Captura de pantalla de EuroMOMO con los datos de España, Portugal y Francia en el período 2019-2022.	64
D.1. Residuos de la deviance del modelo seleccionado para el año 2019 frente a los valores ajustados.	74
D.2. QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2019.	74
D.3. Defunciones observadas y esperadas, a partir de los datos del período 1999-2018 en función de la de las causas de muerte seleccionadas. Galicia, 2019.	76
D.4. Defunciones observadas y esperadas, a partir de los datos del período 1999-2018 en función del capítulo de la causa de muerte. Galicia, 2019.	78
D.5. Residuos de la deviance del modelo seleccionado para el año 2020 frente a los valores ajustados.	81
D.6. QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2020.	81
D.7. Defunciones observadas y esperadas, a partir de los datos del período 1999-2019 en función de la de las causas de muerte seleccionadas. Galicia, 2020.	83
D.8. Defunciones observadas y esperadas, a partir de los datos del período 1999-2019 en función del capítulo de la causa de muerte. Galicia, 2020.	85
D.9. Residuos de la deviance del modelo seleccionado para el año 2021 frente a los valores ajustados.	88
D.10. QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2021.	88
D.11. Defunciones observadas y esperadas, a partir de los datos del período 1999-2020 en función de la de las causas de muerte seleccionadas. Galicia, 2021.	90
D.12. Defunciones observadas y esperadas, a partir de los datos del período 1999-2020 en función del capítulo de la causa de muerte. Galicia, 2021.	92
D.13. Residuos de la deviance del modelo seleccionado para el año 2022 frente a los valores ajustados.	95
D.14. QQ-plot de los residuos de la deviance del modelo seleccionado para el año 2022.	95
D.15. Defunciones observadas y esperadas, a partir de los datos del período 1999-2021 en función de la de las causas de muerte seleccionadas. Galicia, 2022.	97
D.16. Defunciones observadas y esperadas, a partir de los datos del período 1999-2021 en función del capítulo de la causa de muerte. Galicia, 2022.	99