

Máster Universitario en Ciencias Actuariales y Financieras
2023-2024

Trabajo Fin de Máster

“Modelización actuarial de los hábitos
de vida. Escenarios prospectivos de las
tasas de mortalidad”

Ricardo Villanueva Caballero

Tutora

Raquel Pérez Calderón

Madrid, 2024

DETECCIÓN DEL PLAGIO

La Universidad utiliza el programa **Turnitin Feedback Studio** para comparar la originalidad del trabajo entregado por cada estudiante con millones de recursos electrónicos y detecta aquellas partes del texto copiadas y pegadas. Copiar o plagiar en un TFM es considerado una **Falta Grave**, y puede conllevar la expulsión definitiva de la Universidad.



[Incluir en el caso del interés de su publicación en el archivo abierto]

Esta obra se encuentra sujeta a la licencia Creative Commons **Reconocimiento – No Comercial – Sin Obra Derivada**

RESUMEN

El envejecimiento de la población presenta un gran desafío en la actualidad. Por ello, este análisis refleja el efecto de los hábitos de vida saludables y no saludables sobre las tasas de mortalidad futuras, para hombres y mujeres en España. El objetivo que se persigue es concienciar a la población acerca de la repercusión del estilo de vida que tienen, además de hallar los factores que influyen de formas más relevante en las tasas de mortalidad. Para lograrlo, se emplearán los Modelos Lineales Generalizados (GLM) y la modelización de Lee-Carter sobre un determinado número de factores extraídos de distintas encuestas poblacionales.

Palabras clave: Mortalidad, GLM, Lee-Carter

ABSTRACT

The aging population presents a significant challenge today. Therefore, this analysis reflects the effect of healthy and unhealthy lifestyle habits on future mortality rates for men and women in Spain. The objective pursued is to raise awareness among the population about the impact of their lifestyle choices and to identify the factors that most significantly influence mortality rates. To achieve this, Generalized Linear Models (GLM) and Lee-Carter model will be employed on a selected number of factors extracted from various population surveys.

Keywords: Mortality, GLM, Lee-Carter

TABLA DE CONTENIDO

1.	INTRODUCCIÓN.....	1
1.1.	Motivación.....	1
1.2.	Objetivos.....	2
1.3.	Evolución de la esperanza de vida.....	2
1.4.	Dinámicas de envejecimiento y causas	5
1.5.	Tendencias	6
2.	DESCRIPCIÓN DE LA MUESTRA	8
2.1.	Datos.....	8
2.2.	Factores de riesgo	8
3.	METODOLOGÍA.....	54
3.1.	GLM	54
3.2.	Inferencia y contraste de hipótesis en GLM.....	55
3.3.	Modelo de Lee-Carter.....	56
3.4.	Tablas de mortalidad.....	58
4.	RESULTADOS.....	61
4.1.	Resultados GLM.....	61
4.2.	Resultados Lee-Carter	69
5.	PROYECCIONES DE MORTALIDAD: 2020 VS. 2030	77
6.	CONCLUSIONES.....	84
	BIBLIOGRAFÍA	87
	ANEXO	90

ÍNDICE DE FIGURAS

Figura 1.1. <i>Evolución de la Esperanza de Vida 1991-2022</i>	4
Figura 2.1. <i>Porcentaje de fumadores diarios por sexo y grupos de edad, 2020</i>	10
Figura 2.2. <i>Porcentaje de fumadores ocasionales por sexo y grupos de edad, 2020</i>	11
Figura 2.3. <i>Porcentaje de exfumadores por sexo y grupos de edad, 2020</i>	12
Figura 2.4. <i>Porcentaje de no fumadores por sexo y grupos de edad, 2020</i>	12
Figura 2.5. <i>Matriz de correlaciones entre mortalidad femenina y tabaquismo</i>	13
Figura 2.6. <i>Matriz de correlaciones y significatividad entre mortalidad femenina y tabaquismo</i>	14
Figura 2.7. <i>Matriz de correlaciones entre mortalidad masculina y tabaquismo</i>	17
Figura 2.8. <i>Matriz de correlaciones y significatividad entre mortalidad masculina y tabaquismo</i>	18
Figura 2.9. <i>Porcentaje de consumidores de alcohol en los últimos doce meses por sexo y grupos de edad, 2020</i>	19
Figura 2.10. <i>Matriz de correlaciones entre mortalidad femenina y consumo de alcohol en los últimos doce meses</i>	20
Figura 2.11. <i>Matriz de correlaciones y significatividad entre mortalidad femenina y consumo de alcohol en los últimos doce meses</i>	20
Figura 2.12. <i>Matriz de correlaciones entre mortalidad masculina y consumo de alcohol en los últimos doce meses</i>	22
Figura 2.13. <i>Matriz de correlaciones y significatividad entre mortalidad masculina y consumo de alcohol en los últimos doce meses</i>	23
Figura 2.14. <i>Consumo de cannabis en los últimos doce meses para población de 15 a 34 años, 1999-2022</i>	24

Figura 2.15. <i>Consumo de cannabis en los últimos doce meses para población de 35 a 64 años, 1999-2022</i>	24
Figura 2.16. <i>Consumo de cocaína en los últimos doce meses para población de 15 a 34 años, 1999-2022</i>	25
Figura 2.17. <i>Consumo de cocaína en los últimos doce meses para población de 35 a 64 años, 1999-2022</i>	25
Figura 2.18. <i>Consumo de hipnosedantes en los últimos doce meses para población de 15 a 34 años, 1999-2022</i>	26
Figura 2.19. <i>Consumo de hipnosedantes en los últimos doce meses para población de 35 a 64 años, 1999-2022</i>	26
Figura 2.20. <i>Matriz de correlaciones entre mortalidad femenina y consumo de drogas</i>	27
Figura 2.21. <i>Matriz de correlaciones y significatividad entre mortalidad femenina y consumo de drogas</i>	28
Figura 2.22. <i>Matriz de correlaciones entre mortalidad masculina y consumo de drogas</i>	30
Figura 2.23. <i>Matriz de correlaciones y significatividad entre mortalidad masculina y consumo de drogas</i>	31
Figura 2.24. <i>Evolución de la práctica de ejercicio físico, 2011-2022</i>	32
Figura 2.25. <i>Número de días de ejercicio físico semanal para hombres, por grupos de edad</i>	33
Figura 2.26. <i>Número de días de ejercicio físico semanal para mujeres, por grupos de edad</i>	34
Figura 2.27. <i>Matriz de correlaciones entre la mortalidad femenina y el ejercicio semanal</i>	35
Figura 2.28. <i>Matriz de correlaciones y significatividad entre la mortalidad femenina y el ejercicio semanal</i>	35
Figura 2.29. <i>Matriz de correlaciones entre la mortalidad masculina y el ejercicio semanal</i>	36

Figura 2.30. <i>Matriz de correlaciones y significatividad entre la mortalidad masculina y el ejercicio semanal.....</i>	37
Figura 2.31. <i>Índice de Masa Corporal por grupos de edad para hombres, 2020.....</i>	38
Figura 2.32. <i>Índice de Masa Corporal por grupos de edad para mujeres, 2020.....</i>	39
Figura 2.33. <i>Matriz de correlaciones entre la mortalidad femenina y el Índice de Masa Corporal.....</i>	40
Figura 2.34. <i>Matriz de correlaciones y significatividad entre la mortalidad femenina y el Índice de Masa Corporal.....</i>	40
Figura 2.35. <i>Matriz de correlaciones entre la mortalidad masculina y el Índice de Masa Corporal.....</i>	41
Figura 2.36. <i>Matriz de correlaciones y significatividad entre la mortalidad masculina y el Índice de Masa Corporal.....</i>	41
Figura 2.37. <i>Porcentaje de población por horas de sueño y grupos de edad para mujeres, 2020.....</i>	42
Figura 2.38. <i>Porcentaje de población por horas de sueño y grupos de edad para hombres, 2020.....</i>	43
Figura 2.39. <i>Matriz de correlaciones entre la mortalidad femenina y las horas de sueño.....</i>	44
Figura 2.40. <i>Matriz de correlaciones y significatividad entre la mortalidad femenina y las horas de sueño.....</i>	44
Figura 2.41. <i>Matriz de correlaciones entre la mortalidad masculina y las horas de sueño.....</i>	46
Figura 2.42. <i>Matriz de correlaciones y significatividad entre la mortalidad masculina y las horas de sueño.....</i>	47
Figura 2.43. <i>Salarios medios brutos por grupos de edad y decil, 2021.....</i>	48
Figura 2.44. <i>Matriz de correlaciones entre la mortalidad femenina y la renta.....</i>	49
Figura 2.45. <i>Matriz de correlaciones y significatividad entre la mortalidad femenina y la renta.....</i>	49

Figura 2.46. <i>Matriz de correlaciones entre la mortalidad masculina y la renta.....</i>	50
Figura 2.47. <i>Matriz de correlaciones y significatividad entre la mortalidad masculina y la renta.....</i>	50
Figura 2.48. <i>Matriz de correlaciones entre la mortalidad femenina y la edad.....</i>	52
Figura 2.49. <i>Matriz de correlaciones y significatividad entre la mortalidad femenina y la edad.....</i>	52
Figura 2.50. <i>Matriz de correlaciones entre la mortalidad masculina y la edad.....</i>	53
Figura 2.51. <i>Matriz de correlaciones y significatividad entre la mortalidad masculina y la edad.....</i>	53
Figura 4.1. <i>Modelo 1.....</i>	61
Figura 4.2. <i>Modelo 2.....</i>	62
Figura 4.3. <i>Modelo 3.....</i>	63
Figura 4.4. <i>Modelo 4.....</i>	63
Figura 4.5. <i>Modelo 5.....</i>	64
Figura 4.6. <i>Modelo 6.....</i>	65
Figura 4.7. <i>Modelo 7.....</i>	66
Figura 4.8. <i>Modelo 8.....</i>	66
Figura 4.9. <i>Modelo 9.....</i>	67
Figura 4.10. <i>Modelo 10.....</i>	67
Figura 4.11. <i>Resultados de las tasas de mortalidad reales proyectadas para las mujeres.....</i>	70
Figura 4.12. <i>Ajuste Lee-Carter para las tasas de mortalidad reales proyectadas para las mujeres.....</i>	70
Figura 4.13. <i>Proyección de las tasas de mortalidad reales para las mujeres, 2021 – 2030.....</i>	71
Figura 4.14. <i>Resultados de las tasas de mortalidad reales proyectadas para los hombres.....</i>	72

Figura 4.15. <i>Ajuste Lee-Carter para las tasas de mortalidad reales proyectadas para los hombres.....</i>	73
Figura 4.16. <i>Proyección de las tasas de mortalidad reales para los hombres, 2021 – 2030.....</i>	73
Figura 4.17. <i>Resultados del train set.....</i>	74
Figura 4.18. <i>Resultados del test set.....</i>	75
Figura 4.19. <i>Ajuste del backtesting con respecto a las tasas de mortalidad reales.....</i>	75
Figura 5.1. <i>Proyección de las tasas de mortalidad reales y por modelo para los hombres, 2030.....</i>	78
Figura 5.2. <i>Proyección de las tasas de mortalidad reales y por modelo para las mujeres, 2030.....</i>	80

ÍNDICE DE TABLAS

Tabla 2.1. <i>Tasas de tabaquismo y mortalidad</i>	16
Tabla 2.2. <i>Tasas de consumo de alcohol en los últimos doce meses y mortalidad</i>	21
Tabla 2.3. <i>Tasas de consumo de drogas y mortalidad</i>	29
Tabla 2.4. <i>Tasas de mortalidad y edades que superan las horas de sueño recomendadas</i>	45
Tabla 4.1. <i>Predicción de las tasas de mortalidad para cada modelo</i>	68
Tabla 5.1. <i>Proyección de las tasas de mortalidad reales y por modelo para los hombres, 2030</i>	77
Tabla 5.2. <i>Proyección de las tasas de mortalidad reales y por modelo para las mujeres, 2030</i>	79
Tabla 5.3. <i>Diferencias entre las tasas de mortalidad reales y de los modelos para los hombres entre 11 y 64 años</i>	81
Tabla 5.4. <i>Diferencias entre las tasas de mortalidad reales y de los modelos para las mujeres entre 11 y 64 años</i>	82

1. INTRODUCCIÓN

El incremento progresivo de la población anciana a nivel global plantea una serie de desafíos tanto en el ámbito de la salud como en el sociosanitario. Como respuesta a esta situación, se ha generado un impulso en el desarrollo de medidas de promoción de la salud destinadas a contrarrestar los efectos del envejecimiento demográfico, prevenir enfermedades y fomentar el bienestar durante el proceso de envejecimiento. A nivel internacional, diversas entidades se han involucrado en este tema, promoviendo intervenciones que fomenten un envejecimiento activo y saludable. Desde los años 90, la Organización Mundial de la Salud (OMS) ha liderado este esfuerzo, instando a los países miembros a adoptar medidas que promuevan un envejecimiento saludable, subrayando que esto no es simplemente una opción, sino una necesidad apremiante para preservar la salud y mitigar las consecuencias sociales y sanitarias.

1.1. Motivación

Como individuos, estamos inmersos en un mundo donde la salud y la longevidad son aspectos fundamentales que impactan directamente en nuestra calidad de vida y en la de quienes nos rodean. La investigación sobre los factores que afectan a la mortalidad y la longevidad ofrece la oportunidad de identificar los determinantes clave que influyen en la salud y, en última instancia, en la esperanza de vida de las personas. El estudio de los factores de riesgo asociados con la mortalidad y la longevidad nos permite profundizar en la comprensión de los mecanismos biológicos, sociales y ambientales que subyacen a estos fenómenos. Este conocimiento no solo enriquece nuestra comprensión de la salud humana, sino que también puede allanar el camino para el desarrollo de nuevas estrategias de prevención y tratamiento.

Por último, pero no menos importante, está el aspecto personal de esta motivación. Como ser humano, me preocupa mi propia salud y longevidad, así como la de mis seres queridos. Investigar sobre los factores que influyen en estos aspectos vitales no solo me brinda la oportunidad de comprender mejor mi propia salud, sino que también me permite contribuir al bienestar de las personas que me importan y de la sociedad en general.

1.2. Objetivos

Para comprender cómo contribuyen y perjudican a la salud y mortalidad elementos como el estilo de vida, el entorno social y ambiental o ciertos determinantes económicos y culturales, entre otros, se ha de realizar un análisis exhaustivo de cada uno de ellos, así como investigar cuáles son los factores de riesgo más relevantes asociados con enfermedades crónicas y condiciones de salud que afecten a la longevidad y la mortalidad. Dichos factores podrían ser el tabaquismo, la obesidad, la falta de actividad física, la mala alimentación, el estrés e, incluso, el acceso desigual a la atención médica y los recursos de salud.

Por otro lado, se tratará de profundizar en estrategias de prevención y tratamiento de los factores mencionados anteriormente.

Para llevar a cabo este objetivo, se modelizarán varios GLM, uno para cada año de estudio, relacionando un determinado número de variables saludables (como la actividad física, las horas de sueño, etc.) y no saludables (como el tabaquismo, el alcoholismo, el consumo de drogas...) con la mortalidad. Con esto, se logrará saber qué factores de riesgo son relevantes en el estudio de la mortalidad y, a partir de ahí, se empleará el modelo de Lee-Carter para estimar la mortalidad futura, en base a la información calculada previamente. Finalmente, se compararán las tasas de mortalidad de la base de datos con las calculadas a través del modelo, para comprobar si los factores de riesgo realmente influyen en la mortalidad.

1.3. Evolución de la esperanza de vida

A lo largo de la historia, la esperanza de vida ha sido un indicador revelador de las condiciones socioeconómicas, ambientales y de salud en cada período. Desde tiempos antiguos hasta la actualidad, hemos sido testigos de una notable evolución en este aspecto, influenciada por una amplia gama de factores.

En las civilizaciones antiguas, la esperanza de vida era notablemente reducida, generalmente oscilando entre los 20 y 30 años. Este fenómeno se atribuye principalmente a las adversas condiciones de vida, la escasez de recursos, la limitada disponibilidad de alimentos nutritivos y agua potable, así como a la prevalencia de enfermedades infecciosas y epidemias. Las personas estaban expuestas a diversas enfermedades

mortales, y las tasas de mortalidad infantil eran elevadas, ejerciendo un impacto significativo en la esperanza de vida media.

Durante la Edad Media y el Renacimiento, aunque se registraron algunos avances en términos de higiene y conocimientos médicos, la esperanza de vida aún permanecía relativamente baja. Las epidemias de enfermedades como la peste bubónica o peste negra y la tuberculosis seguían siendo comunes y mortales. Además, las condiciones de vida eran duras para la mayoría, con una nutrición deficiente y una atención médica insuficiente. En esta época, la esperanza de vida media se situaba generalmente en el rango de los 30 a 40 años.

Posteriormente, a partir de la segunda mitad del siglo XVIII, la Revolución Industrial provocó cambios significativos en la sociedad y la economía, aunque inicialmente no se tradujo en un impacto positivo inmediato en la esperanza de vida. A pesar de mejoras en la tecnología y la producción, las condiciones laborales en las fábricas y minas eran peligrosas y poco saludables, lo que resultaba en lesiones graves y enfermedades profesionales. La rápida urbanización también contribuyó a la propagación de enfermedades. No obstante, hacia finales del siglo XIX y principios del XX, se observaron mejoras en la salud pública, la higiene y la atención médica, lo que condujo a un incremento gradual en la esperanza de vida, impulsado principalmente por una reducción aproximada del 40% en las tasas de mortalidad infantil.

El siglo XX presencié avances revolucionarios en medicina y tecnología médica. Se introdujeron vacunas, antibióticos y tratamientos médicos más efectivos, lo que redujo significativamente la incidencia de enfermedades infecciosas y aumentó la esperanza de vida en muchas partes del mundo. Además, mejoras en la nutrición, la educación sanitaria y las condiciones de vida también jugaron un papel crucial en este aumento. La esperanza de vida media superó los 70 años en muchos países desarrollados hacia finales de este siglo.

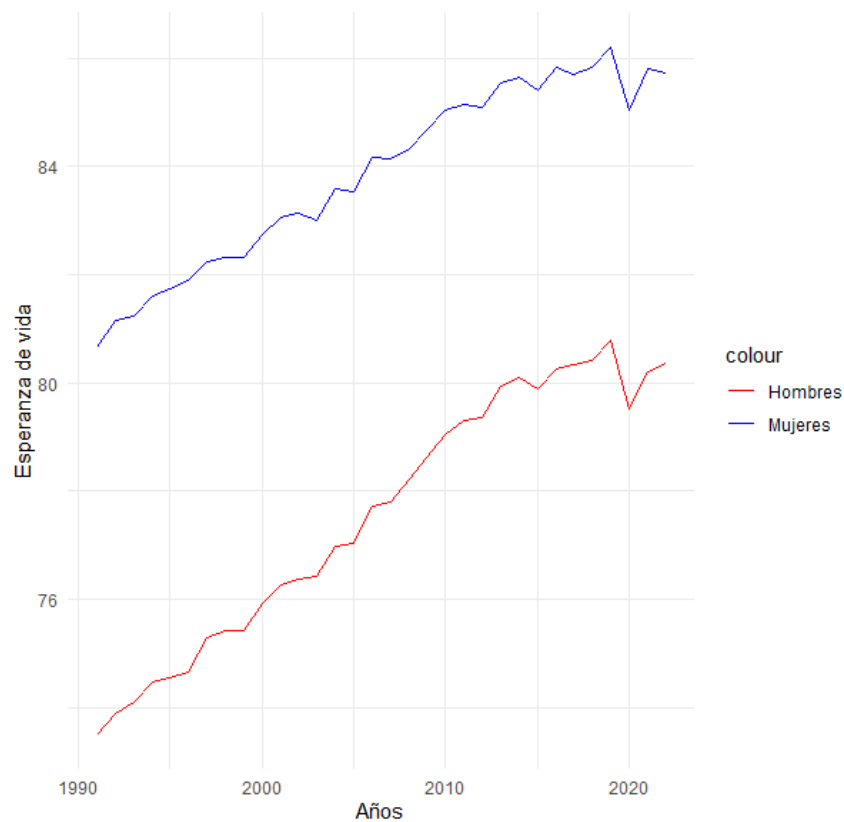
En el siglo XXI, la esperanza de vida ha continuado aumentando gracias a los avances continuos en medicina, investigación científica y conciencia sobre la salud. Se han logrado progresos significativos en la prevención, diagnóstico y tratamiento de enfermedades, contribuyendo a un envejecimiento más saludable de la población en muchos países. Además, los avances en genética y medicina regenerativa están abriendo nuevas posibilidades para prolongar la vida y mejorar la calidad de vida en la vejez. Sin embargo, persisten desafíos como el aumento de enfermedades crónicas relacionadas con

el estilo de vida y la disparidad en el acceso a la atención médica, que requieren un enfoque continuo en la promoción de la salud y la equidad en el cuidado de la salud.

A continuación, se muestra un gráfico con datos extraídos de los Indicadores demográficos básicos del INE, donde se puede observar claramente que las mujeres tienen una esperanza de vida superior a los hombres, sobre todo al principio de la serie de tiempo.

Figura 1.1

Evolución de la Esperanza de Vida 1991-2022



Fuente: Elaboración propia

La figura 1.1, de elaboración propia a través de R, muestra la evolución de la esperanza de vida entre los años 1991 y 2022 en España.

Como se venía comentando, durante el siglo XX, la esperanza de vida llegó a sobrepasar la edad de 70 años en los países más desarrollados, como sería el caso de España. Su población se duplicó, pasó de 18,6 millones de habitantes a 39,5 millones, y las tasas de mortalidad disminuyeron del 2,83% al 0,92% durante ese período (Gisbert & Pallejà,

2006). Este fenómeno podría explicarse por el aumento de la esperanza de vida, que sufrió un crecimiento para todas las edades a lo largo de los 1900s, a pesar de haber sufrido conflictos armados como la Guerra de Marruecos o Guerra del Rif (1911-1927) y la Guerra Civil Española (1936-1939).

Por otro lado, ya entrando en el siglo XXI, la esperanza de vida siguió aumentando, con un patrón de evolución muy similar entre hombres y mujeres, aunque cabe destacar que hay una diferencia entre ambos de 6 años aproximadamente. Para el año 2020, la esperanza de vida percibió una disminución considerable para ambos sexos, posiblemente causada por la pandemia de Covid-19, que fue responsable de una gran cantidad de muertes en el país. España fue uno de los países más afectados de Europa por este virus, aumentando en gran medida sus tasas de mortalidad para los años 2020 y 2021, ya que no sólo afectó de forma directa a dicha mortalidad, sino que también colapsó los sistemas de salud y se interrumpieron tratamientos médicos para centrar el foco sanitario en la mitigación de la pandemia.

1.4. Dinámicas de envejecimiento y causas

Cuando se aborda el tema del envejecimiento, se hace referencia al proceso natural e inevitable que experimentan todos los seres humanos a lo largo de sus vidas. A pesar de ser un fenómeno arraigado en la experiencia humana, las dinámicas del envejecimiento están experimentando cambios significativos, motivados por una variedad de factores, entre los que destacan los avances médicos, los cambios en el estilo de vida y las transformaciones en la estructura demográfica de la población.

Como se mencionó previamente, la esperanza de vida ha experimentado un aumento constante a lo largo de la historia. Los avances médicos han posibilitado la investigación, creación y desarrollo de tratamientos efectivos frente a una amplia gama de enfermedades, además de contribuir a su prevención y detección temprana. La disminución de las tasas de mortalidad infantil en épocas modernas ha sido facilitada por mejoras en la atención médica prenatal y neonatal, así como por la disponibilidad de vacunas y medidas de saneamiento básico, que han mejorado las condiciones de vida y la preparación para hacer frente a posibles enfermedades futuras.

Este aumento en la longevidad, junto con la disminución de las tasas de natalidad, está impulsando el envejecimiento demográfico, un fenómeno global que está transformando

de manera inevitable la estructura poblacional en numerosos países del mundo. Este cambio puede acarrear implicaciones socioeconómicas y de salud, como una posible reducción de la mano de obra joven, desafíos en la asignación de recursos en los sistemas de pensiones y una mayor demanda de servicios de salud.

Asimismo, el proceso de envejecimiento conlleva la persistencia de enfermedades crónicas, problemas cardiovasculares, discapacidades físicas y deterioro cognitivo. Sin embargo, estos efectos pueden ser mitigados mediante la adopción de hábitos de vida saludables, como una alimentación equilibrada, la práctica regular de ejercicio físico y la abstención de consumir sustancias nocivas para la salud.

1.5. Tendencias

Las tendencias de mortalidad y esperanza de vida en España han experimentado transformaciones notables en las últimas décadas. De manera general, la esperanza de vida ha experimentado un incremento constante, mientras que las tasas de mortalidad han decrecido, gracias a los avances en la medicina, las mejoras en las condiciones de vida y los programas de salud pública. No obstante, al igual que muchos países desarrollados, España se enfrenta a desafíos demográficos, incluido el envejecimiento de la población y sus implicaciones para la sostenibilidad de los sistemas de pensiones y salud.

En términos generales, la esperanza de vida al nacer en España ha registrado un aumento progresivo en las últimas décadas, situándose en aproximadamente 80 años para los hombres y 86 años para las mujeres. Se prevé que este indicador continúe en ascenso en España en los próximos años, aunque posiblemente a un ritmo más moderado. Factores como los avances médicos, el acceso a la atención sanitaria y cambios en el estilo de vida, como los hábitos saludables que se discutirán más adelante, podrían seguir contribuyendo a este crecimiento más lento, influenciado por el envejecimiento de la población y los efectos residuales de la pandemia de Covid-19. Paralelamente, España enfrenta otro desafío demográfico, caracterizado por el descenso en las tasas de natalidad.

Por otra parte, en lo que respecta a la mortalidad, se observa una tendencia descendente constante en las tasas de mortalidad en general y en las principales causas de muerte, como las enfermedades cardiovasculares y el cáncer, debido a mejoras en la atención médica, estilos de vida más saludables y programas de prevención.

Es crucial tener presente que estas tendencias pueden estar condicionadas por factores socioeconómicos, acceso a la atención médica y otros determinantes de la salud. Además, la pandemia de COVID-19 ha tenido un impacto significativo en la mortalidad y podría influir en las tendencias a corto plazo.

2. DESCRIPCIÓN DE LA MUESTRA

2.1. Datos

En este estudio se han utilizado las bases de datos tanto del Human Mortality Database (HMD), como del Instituto Nacional de Estadística (INE).

En cuanto a los datos procedentes del HMD, se han descargado el número de muertes, nacimientos, tamaño de población, población expuesta a factores de riesgo (por período y cohorte), tasas de mortalidad (por período y cohorte) y tablas de mortalidad (por período y cohorte), para España y en el período comprendido entre los años 1908 y 2021. Las tablas de mortalidad son herramientas estadísticas que muestran la probabilidad de fallecimiento, es decir, proporcionan información valiosa sobre la salud y el bienestar de una población. Por ello, para este análisis, serán los únicos datos utilizados de entre los descargados en el HMD, pues el objetivo es estudiar cómo afectan determinados factores de riesgo en la mortalidad. Dichos factores serán presentados más adelante. Con el objetivo de simplificar el estudio, se han utilizado las tablas de mortalidad del período 1960-2021, haciendo distinción entre hombres y mujeres, por lo que en el análisis se podrá distinguir entre ambos sexos.

2.2. Factores de riesgo

En cuanto a la información del INE, se han utilizado los registros históricos de la Encuesta Nacional de Salud de España y de la Encuesta Europea de Salud en España, desde el año 1987 hasta el 2021, para obtener los factores de riesgo que perjudican y benefician a la salud. Para simplificar el análisis, se han utilizado los datos más recientes de cada encuesta y estudio, es decir, los pertenecientes a 2020 y 2021. De este modo, se han seguido rigurosamente las indicaciones de la literatura empleada para el estudio, siendo así que, por un lado, los posibles factores de riesgo que afectan a la mortalidad serían los siguientes:

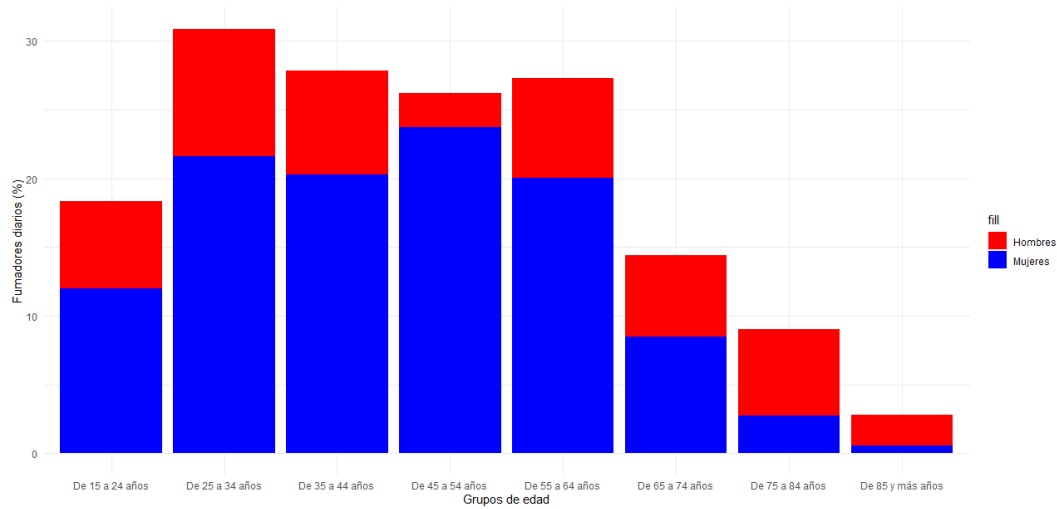
1. Tabaquismo,
2. Alcoholismo,
3. Consumo de drogas,
4. Actividad física,
5. Índice de masa corporal,

6. Horas de sueño,
7. Nivel de renta y
8. Edad

El primer factor es el tabaquismo. Este factor de riesgo afecta negativamente a la longevidad. Según el Plan Nacional sobre Drogas del Gobierno de España, en el país fallecen anualmente más de cincuenta mil personas a causa del consumo de tabaco. Esta cifra supera a la suma del número de fallecidos por accidente de tráfico y el del consumo de drogas ilegales. El tabaco es el precedente de enfermedades como cáncer de pulmón, hipertensión arterial, cáncer de laringe, enfermedades coronarias (como anginas o infartos de miocardio) o problemas cerebrovasculares (como trombosis, hemorragias o embolias) (Observatorio Español de las Drogas y las Adicciones, 2023). Según la OCDE en su artículo “Health at a Glance 2023”, “el tabaquismo es la principal causa de multitud de enfermedades. La OMS estima que fumar tabaco mata a ocho millones de personas al año en todo el mundo. Más de un millón doscientas mil de estas muertes se deben a respirar el humo de otra persona y sesenta y cinco mil se producen entre niños” (OECD, 2023). En cuanto a los datos descargados, el INE estima a través de la Encuesta Nacional de Salud y de la Encuesta Europea de Salud en España la tasa de tabaquismo en base a variables como el consumo en personas mayores de 15 años de edad, el tipo de consumidor, o la frecuencia del consumo, entre otras. Se muestra a continuación la figura 2.1 sobre el consumo de tabaco diario, dividido por grupos de edad.

Figura 2.1

Porcentaje de fumadores diarios por sexo y grupos de edad, 2020

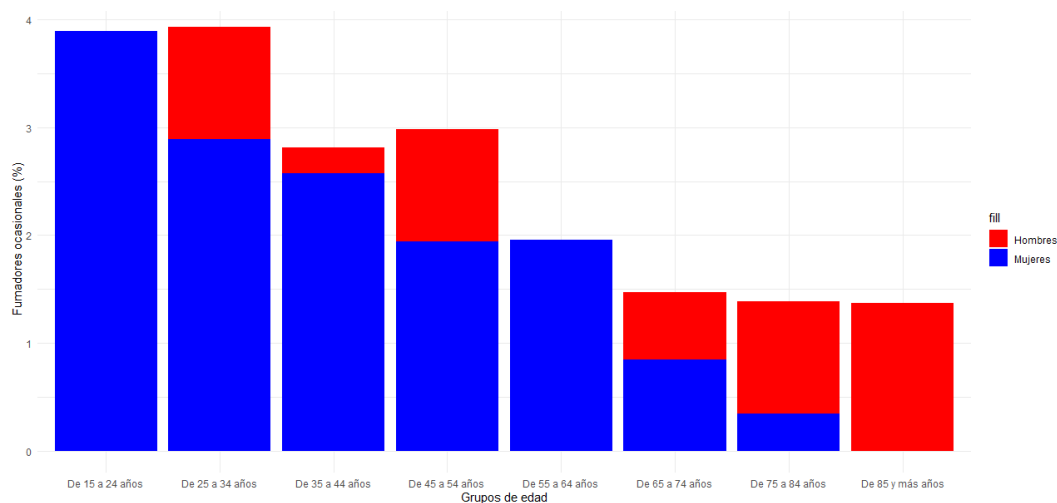


Fuente: Elaboración propia

Como podemos ver en la figura 2.1 sobre los fumadores diarios, en general, los hombres consumen más cigarrillos al día que las mujeres. La mayor diferencia se da entre los 25 y los 34 años de edad, donde aproximadamente un 31% de la población masculina fuma, a diferencia de la femenina, en la que fuma alrededor del 22%. Sin embargo, el grupo de edad femenino que más consume tabaco diario es el comprendido entre los 45 y los 54 años, con cerca del 24% de fumadoras. Los porcentajes mínimos para ambos sexos ocurren a partir de los 85 años.

Figura 2.2

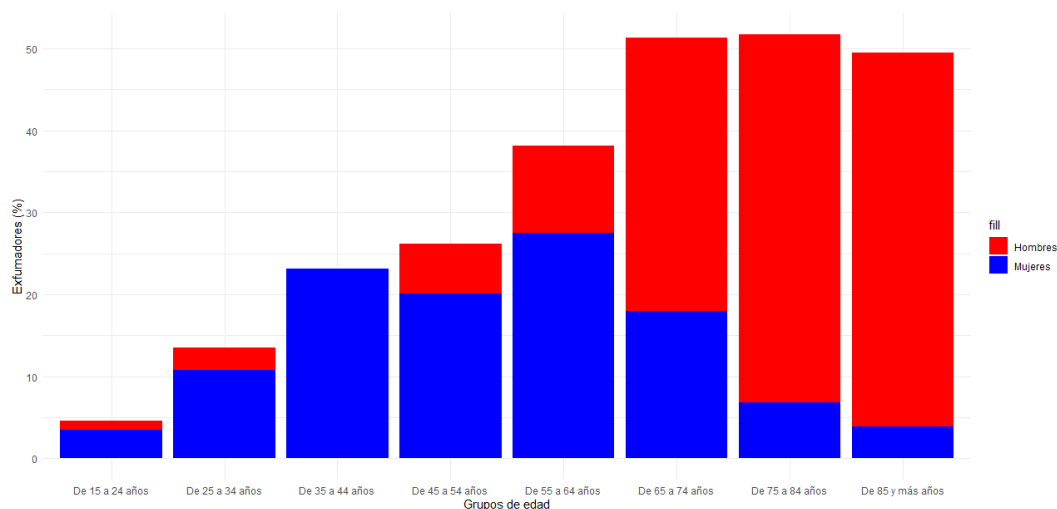
Porcentaje de fumadores ocasionales por sexo y grupos de edad, 2020



Fuente: Elaboración propia

En cuanto a los fumadores ocasionales, claramente se aprecia un porcentaje mucho menor para ambos sexos que los consumidores diarios, pues la escala de estos estaba comprendida entre el 0% y el 40% de la población, mientras que los ocasionales están acotados entre el 0% y el 4% únicamente. Este dato no es necesariamente positivo, ya que el mayor porcentaje de población consume de tabaco diariamente. En la figura 2.2 destaca que las mujeres consumen más tabaco que los hombres de forma ocasional para los grupos de edad comprendidos entre los 15 y 24 años (cerca del 4%) y entre los 55 y los 64 (alrededor del 2%). Curiosamente, las mujeres no consumen tabaco a partir de los 85 años. Este gráfico muestra un claro abandono paulatino del consumo de tabaco para ambos sexos, aunque los hombres perciben un ligero repunte entre los 45 y los 54 años.

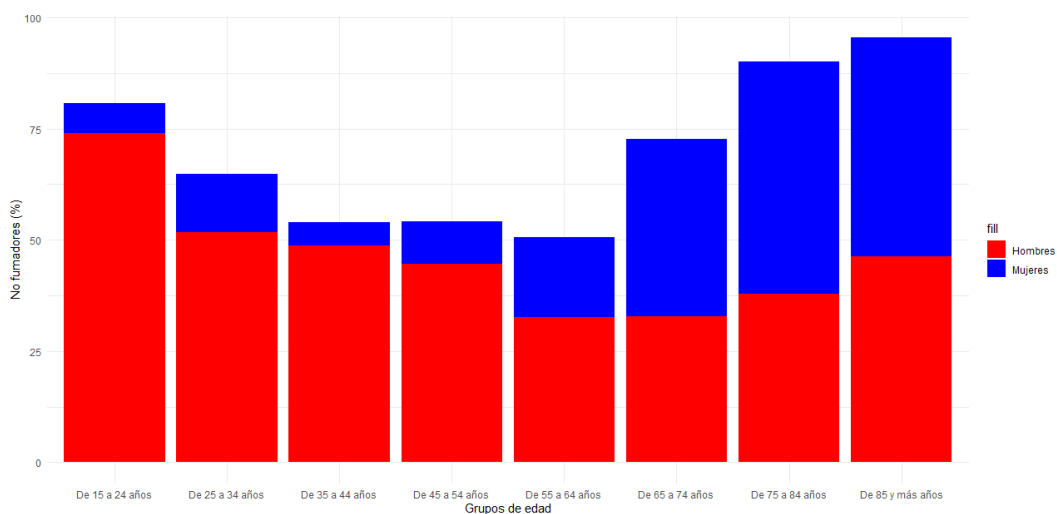
Figura 2.3
Porcentaje de exfumadores por sexo y grupos de edad, 2020



Fuente: Elaboración propia

Como era de esperar, en la figura 2.3, las mayores tasas de exfumadores se dan entre los hombres, pues han sido los que han mostrado mayor consumo de tabaco tanto diariamente, como de forma ocasional. No obstante, entre los 35 y los 44 años, el porcentaje de mujeres que abandonaron el tabaquismo fue, aproximadamente, un 3% superior que el de hombres. Además, se puede ver cómo la tasa de exfumadoras es muy baja a partir de los 85 años porque, como se puede recordar, la población de fumadoras diarias para esas edades era ínfima (menor al 1%) y la de fumadoras ocasionales, 0%.

Figura 2.4
Porcentaje de no fumadores por sexo y grupos de edad, 2020



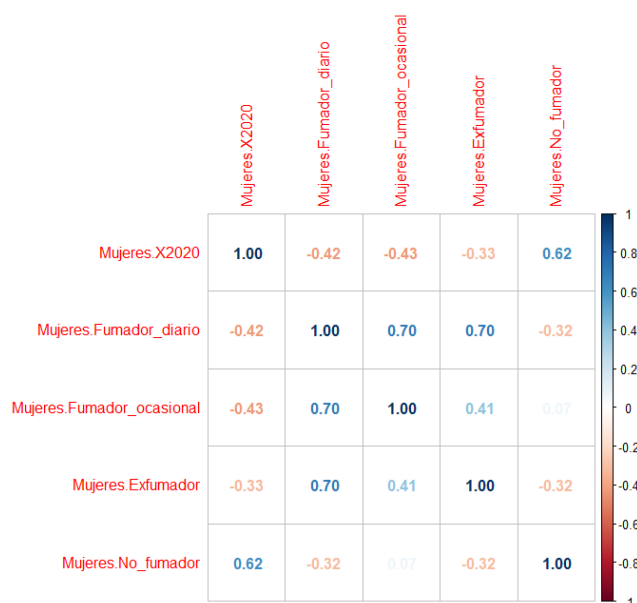
Fuente: Elaboración propia

En el caso de los no fumadores, en la figura 2.4, la población femenina muestra unas mayores tasas de no consumo. Las mayores diferencias se dan en los grupos de edad más longevos, mientras que las menores diferencias entre los 35 y los 44 años.

Las figuras 2.5 y 2.6 muestran las correlaciones entre la mortalidad de 2020 para las mujeres y los diferentes factores relacionados con el tabaquismo.

La primera de ellas es una matriz de correlaciones simple, que muestra la correlación lineal entre las variables *mortalidad*, *fumador diario*, *fumador ocasional*, *exfumador* y *no fumador*. Como es bien sabido, el coeficiente de correlación lineal abarca los valores situados entre -1 y 1. Con el objetivo de hacer la matriz más visual, además, se clasifican los coeficientes en una escala cromática, asociando a la correlación positiva máxima el color azul oscuro y a la negativa mínima, el rojo oscuro.

Figura 2.5
Matriz de correlaciones entre mortalidad femenina y tabaquismo



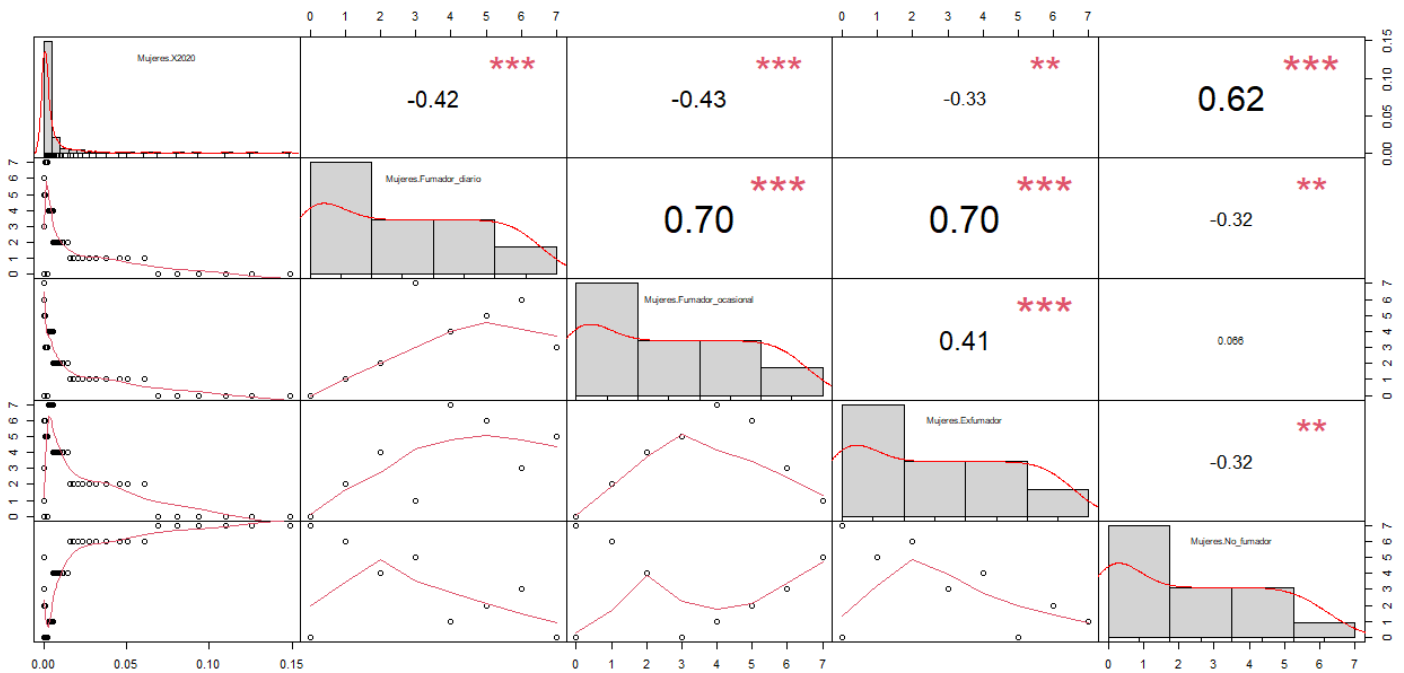
Fuente: Elaboración propia

La matriz descrita anteriormente tiene una limitación, y es que sólo proporciona información sobre la relación lineal que existe entre las variables, pero no muestra si los factores de riesgo realmente tienen efectos significativos sobre la mortalidad. Por ello, se ha extraído una segunda tabla, la figura 2.6, que es una matriz de correlaciones lineales similar a la primera, pero en la que se muestra la distribución de cada una de las variables

y lo más interesante, el nivel de significatividad de sus efectos sobre la mortalidad. El número de asteriscos está relacionado con el nivel de confianza necesario para asumir la influencia de los factores en la mortalidad, es decir, tres asteriscos implican un nivel de confianza del 99%, dos asteriscos implican una confianza del 95% y un asterisco, 90%. En caso de que al coeficiente obtenido no se le asigne ningún asterisco, implicaría que dicha variable no es estadísticamente significativa.

Figura 2.6

Matriz de correlaciones y significatividad entre mortalidad femenina y tabaquismo



Fuente: Elaboración propia

En cuanto a los resultados de las matrices, se tiene la correlación lineal entre las tasas de mortalidad de las mujeres en el año 2020 y los factores relacionados con el tabaquismo, es decir, las tasas de fumadores diarios, fumadores ocasionales, exfumadores y no fumadores. Todos los factores son estadísticamente significativos y tienen influencia en las tasas de mortalidad, resultando una correlación moderada negativa en los tres primeros y fuerte positiva en las tasas de no fumadores. Esto quiere decir que, cuanto mayor sea el número de no fumadores, mayor es el número de defunciones y que, cuanto mayores sean las tasas de población que consume tabaco, menores son las tasas de mortalidad. A priori, estos resultados pueden parecer contrarios a los que se pensarían de forma intuitiva, no obstante, se va a mostrar una explicación del comportamiento con los datos utilizados.

La tabla 2.1, expuesta a continuación, contiene las tasas de tabaquismo de los diferentes factores de riesgo, así como las tasas de mortalidad asociadas a cada factor y a cada edad, para las mujeres.

Tabla 2.1
Tasas de tabaquismo y mortalidad

Edad	Mortalidad_2020	Fumador_diario	Fumador_ocasional	Exfumador	No_fumador
0	0,0027	0	0	0	0
1	0,00023	0	0	0	0
2	0,00012	0	0	0	0
3	0,00012	0	0	0	0
4	0,00008	0	0	0	0
5	0,00008	0	0	0	0
6	0,00006	0	0	0	0
7	0,00005	0	0	0	0
8	0,00004	0	0	0	0
9	0,00004	0	0	0	0
10	0,00006	0	0	0	0
11	0,00009	0	0	0	0
12	0,00006	0	0	0	0
13	0,0001	0	0	0	0
14	0,0001	0	0	0	0
15	0,0001	0,1832	0,0315	0,0462	0,739
16	0,00014	0,1832	0,0315	0,0462	0,739
17	0,00017	0,1832	0,0315	0,0462	0,739
18	0,00024	0,1832	0,0315	0,0462	0,739
19	0,00032	0,1832	0,0315	0,0462	0,739
20	0,00038	0,1832	0,0315	0,0462	0,739
21	0,00034	0,1832	0,0315	0,0462	0,739
22	0,00034	0,1832	0,0315	0,0462	0,739
23	0,00043	0,1832	0,0315	0,0462	0,739
24	0,00031	0,1832	0,0315	0,0462	0,739
25	0,00051	0,3089	0,0393	0,1348	0,517
26	0,00038	0,3089	0,0393	0,1348	0,517
27	0,00043	0,3089	0,0393	0,1348	0,517
28	0,00042	0,3089	0,0393	0,1348	0,517
29	0,00044	0,3089	0,0393	0,1348	0,517
30	0,00055	0,3089	0,0393	0,1348	0,517
31	0,00058	0,3089	0,0393	0,1348	0,517
32	0,00062	0,3089	0,0393	0,1348	0,517
33	0,00057	0,3089	0,0393	0,1348	0,517
34	0,0006	0,3089	0,0393	0,1348	0,517
35	0,00055	0,2783	0,0281	0,2058	0,488
36	0,00065	0,2783	0,0281	0,2058	0,488
37	0,00073	0,2783	0,0281	0,2058	0,488
38	0,00076	0,2783	0,0281	0,2058	0,488
39	0,00094	0,2783	0,0281	0,2058	0,488
40	0,00093	0,2783	0,0281	0,2058	0,488
41	0,00101	0,2783	0,0281	0,2058	0,488
42	0,00107	0,2783	0,0281	0,2058	0,488
43	0,00123	0,2783	0,0281	0,2058	0,488
44	0,00131	0,2783	0,0281	0,2058	0,488
45	0,00155	0,2623	0,0298	0,2614	0,447
46	0,0016	0,2623	0,0298	0,2614	0,447
47	0,00192	0,2623	0,0298	0,2614	0,447
48	0,00225	0,2623	0,0298	0,2614	0,447
49	0,00246	0,2623	0,0298	0,2614	0,447
50	0,0028	0,2623	0,0298	0,2614	0,447
51	0,00318	0,2623	0,0298	0,2614	0,447
52	0,0037	0,2623	0,0298	0,2614	0,447
53	0,00435	0,2623	0,0298	0,2614	0,447
54	0,00462	0,2623	0,0298	0,2614	0,447
55	0,00522	0,2733	0,0194	0,3817	0,326
56	0,00569	0,2733	0,0194	0,3817	0,326
57	0,0062	0,2733	0,0194	0,3817	0,326
58	0,00692	0,2733	0,0194	0,3817	0,326
59	0,00749	0,2733	0,0194	0,3817	0,326
60	0,0083	0,2733	0,0194	0,3817	0,326
61	0,00934	0,2733	0,0194	0,3817	0,326
62	0,01029	0,2733	0,0194	0,3817	0,326
63	0,0113	0,2733	0,0194	0,3817	0,326
64	0,01243	0,2733	0,0194	0,3817	0,326
65	0,01356	0,144	0,0147	0,513	0,328
66	0,0143	0,144	0,0147	0,513	0,328
67	0,01577	0,144	0,0147	0,513	0,328
68	0,01704	0,144	0,0147	0,513	0,328
69	0,01868	0,144	0,0147	0,513	0,328
70	0,02062	0,144	0,0147	0,513	0,328
71	0,02199	0,144	0,0147	0,513	0,328
72	0,02443	0,144	0,0147	0,513	0,328
73	0,02681	0,144	0,0147	0,513	0,328
74	0,0301	0,144	0,0147	0,513	0,328
75	0,03323	0,0905	0,0139	0,5176	0,378
76	0,03652	0,0905	0,0139	0,5176	0,378
77	0,04123	0,0905	0,0139	0,5176	0,378
78	0,04502	0,0905	0,0139	0,5176	0,378
79	0,0501	0,0905	0,0139	0,5176	0,378
80	0,05612	0,0905	0,0139	0,5176	0,378
81	0,06529	0,0905	0,0139	0,5176	0,378
82	0,07188	0,0905	0,0139	0,5176	0,378
83	0,07961	0,0905	0,0139	0,5176	0,378
84	0,0894	0,0905	0,0139	0,5176	0,378
85	0,10193	0,028	0,0137	0,4947	0,464
86	0,11382	0,028	0,0137	0,4947	0,464
87	0,12667	0,028	0,0137	0,4947	0,464
88	0,14277	0,028	0,0137	0,4947	0,464
89	0,16048	0,028	0,0137	0,4947	0,464
90	0,17893	0,028	0,0137	0,4947	0,464

Fuente: Elaboración propia

Como se puede observar, las tasas de tabaquismo diario alcanzan su valor máximo entre los 25 y los 34 años y, a partir de ahí, comienzan a descender de forma gradual hasta llegar al tramo de edades entre los 65 y los 90 años, donde caen de forma más repentina. A su vez, las tasas de mortalidad van aumentando a medida que se incrementa la edad. Esto podría explicar una posible confusión en la comprensión de los coeficientes devueltos. Del mismo modo ocurre con las tasas de fumadores ocasionales, exfumadores y no fumadores.

Esta explicación es análoga al resto de matrices de correlaciones del documento. Sin embargo, se explicará en detalle cada una de ellas en sus respectivos apartados.

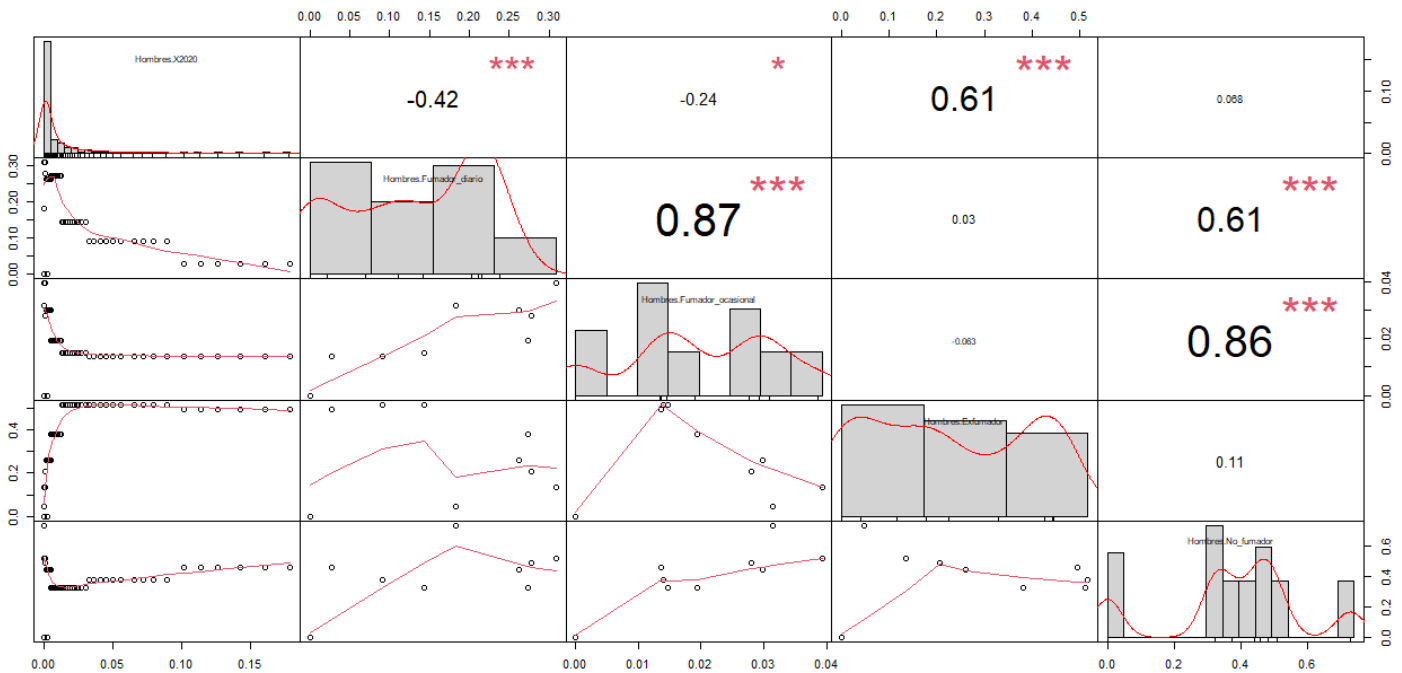
Figura 2.7
Matriz de correlaciones entre mortalidad masculina y tabaquismo



Fuente: Elaboración propia

Figura 2.8

Matriz de correlaciones y significatividad entre mortalidad masculina y tabaquismo



Fuente: Elaboración propia

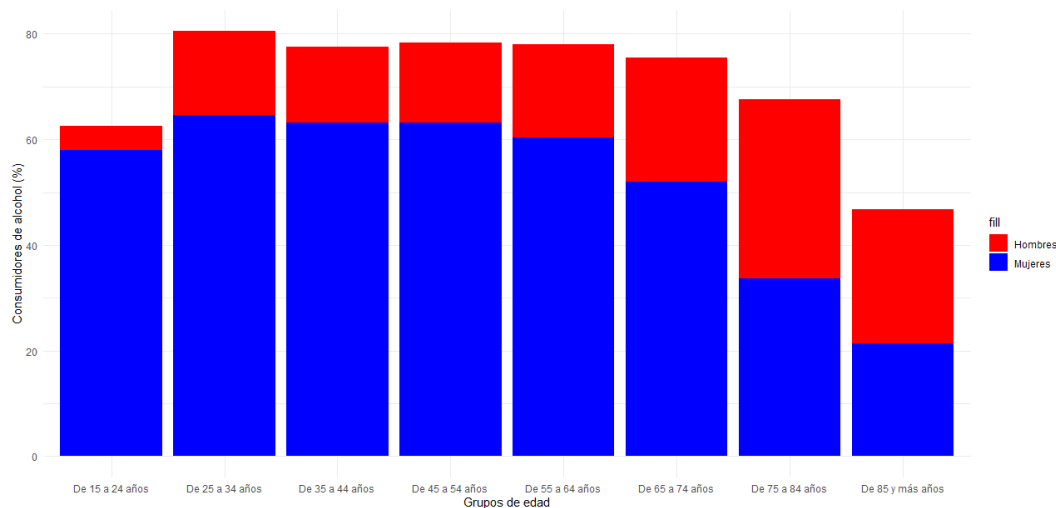
Las figuras 2.7 y 2.8 muestran las correlaciones entre las tasas de mortalidad para los hombres del año 2020 y las tasas de tabaquismo. A diferencia de lo que ocurre con las mujeres, las tasas de no fumadores no son significativas para explicar la mortalidad. Cabe destacar que las tasas de fumadores ocasionales sólo son significativas con un nivel de confianza del 90%. Algo que también puede llamar la atención es el signo del coeficiente de las tasas de exfumadores, que se puede explicar por la no significatividad de las tasas de no fumadores, ya que, si se asume entonces que toda la población masculina ha fumado en algún momento, es lógico pensar que, en algún momento, gran parte de esa población que consume tabaco dejará de hacerlo. Por tanto, a medida que evoluciona la variable edad, aumentará la tasa de población que ya no fuma.

El segundo factor analizado es el alcoholismo. También afecta de forma negativa a la longevidad. El “Plan Nacional sobre Drogas” del Gobierno de España destaca, entre otras, hipertensión arterial, cirrosis hepática, cardiopatías, demencia o psicosis como enfermedades crónicas como consecuencia del consumo excesivo de alcohol (Observatorio Español de las Drogas y las Adicciones, 2023). La tasa de alcoholismo

calculada en la Encuesta Nacional de Salud se basa en la frecuencia de consumo de alcohol, la edad de inicio de consumo y la cantidad ingerida.

Figura 2.9

Porcentaje de consumidores de alcohol en los últimos doce meses por sexo y grupos de edad, 2020



Fuente: Elaboración propia

La figura 2.9 muestra el porcentaje de población que ha consumido alcohol en los últimos doce meses, distinguiendo entre hombres y mujeres. En general, se aprecia que las mujeres consumen menos alcohol que los hombres y que, para ambos sexos, el grupo de edad con un consumo mayor es el de 25 a 34 años, con un 80% de hombres y 65% de mujeres. Ambos sexos mantienen el consumo entre los 25 y los 64 años, momento en el que comienza un suave descenso del consumo para hombres y algo más pronunciado para mujeres.

Las figuras 2.10 y 2.11 muestran la correlación lineal entre la mortalidad de 2020 de las mujeres y las tasas de alcoholismo.

Figura 2.10

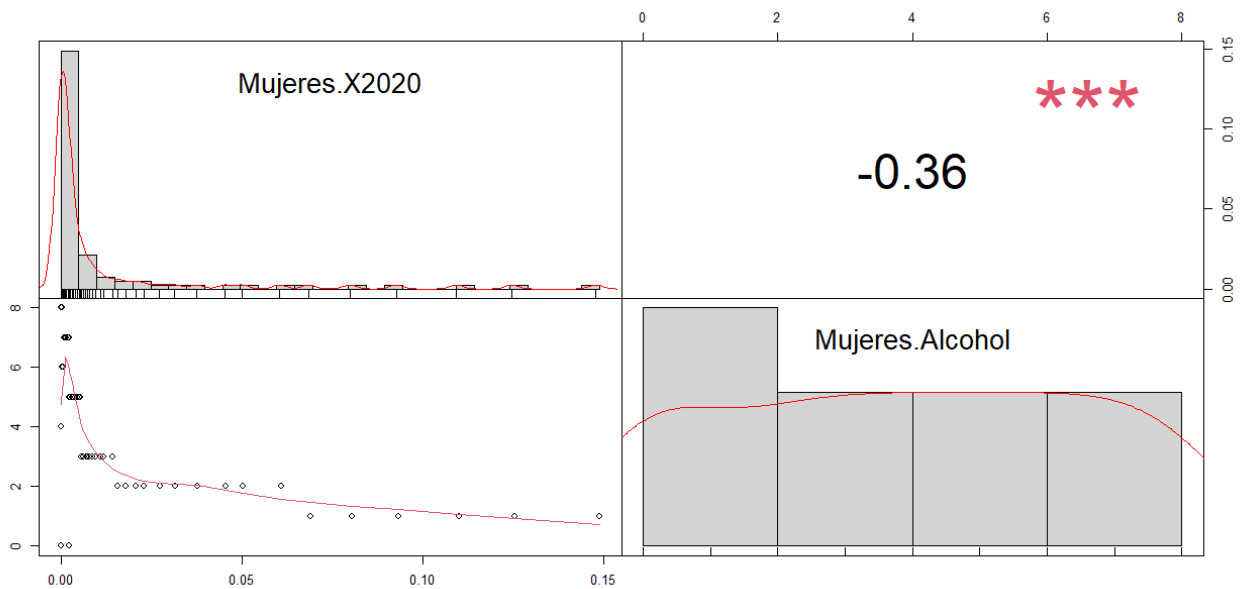
Matriz de correlaciones entre mortalidad femenina y consumo de alcohol en los últimos doce meses



Fuente: Elaboración propia

Figura 2.11

Matriz de correlaciones y significatividad entre mortalidad femenina y consumo de alcohol en los últimos doce meses



Fuente: Elaboración propia

Del mismo modo que ocurría con las tasas de tabaquismo, el coeficiente resultante puede parecer que no tiene sentido. No obstante, a continuación, se procede a dar una explicación del comportamiento de los datos.

Tabla 2.2*Tasas de consumo de alcohol en los últimos doce meses y mortalidad*

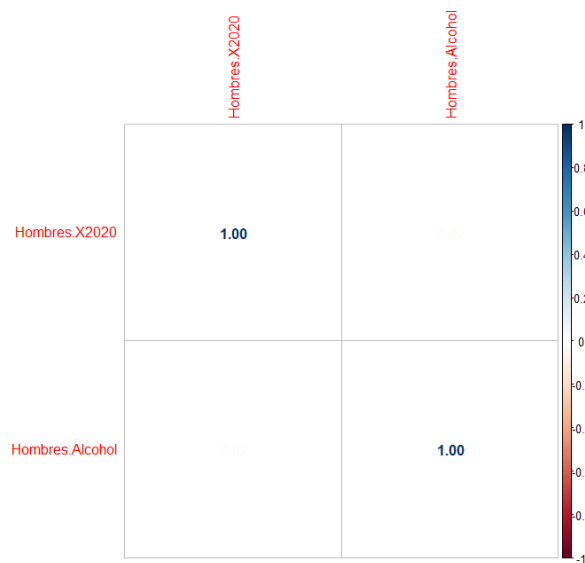
Edad	Mortalidad_2020	Alcohol
0	0,00228	0
1	0,00017	0
2	0,00008	0
3	0,00009	0
4	0,00002	0
5	0,00005	0
6	0,00006	0
7	0,00008	0
8	0,00008	0
9	0,00004	0
10	0,00005	0
11	0,00007	0
12	0,00006	0
13	0,00005	0
14	0,00010	0
15	0,00011	0,5802
16	0,00011	0,5802
17	0,00013	0,5802
18	0,00012	0,5802
19	0,00011	0,5802
20	0,00012	0,5802
21	0,00011	0,5802
22	0,00012	0,5802
23	0,00014	0,5802
24	0,00016	0,5802
25	0,00017	0,6466
26	0,00019	0,6466
27	0,00020	0,6466
28	0,00020	0,6466
29	0,00019	0,6466
30	0,00022	0,6466
31	0,00026	0,6466
32	0,00029	0,6466
33	0,00029	0,6466
34	0,00026	0,6466
35	0,00033	0,6317
36	0,00039	0,6317
37	0,00045	0,6317
38	0,00042	0,6317
39	0,00044	0,6317
40	0,00059	0,6317
41	0,00052	0,6317
42	0,00067	0,6317
43	0,00071	0,6317
44	0,00077	0,6317
45	0,00090	0,632
46	0,00095	0,632
47	0,00107	0,632
48	0,00119	0,632
49	0,00139	0,632
50	0,00151	0,632
51	0,00158	0,632
52	0,00200	0,632
53	0,00207	0,632
54	0,00225	0,632
55	0,00249	0,604
56	0,00277	0,604
57	0,00313	0,604
58	0,00345	0,604
59	0,00355	0,604
60	0,00405	0,604
61	0,00415	0,604
62	0,00453	0,604
63	0,00515	0,604
64	0,00546	0,604
65	0,00581	0,52
66	0,00633	0,52
67	0,00699	0,52
68	0,00729	0,52
69	0,00788	0,52
70	0,00888	0,52
71	0,00962	0,52
72	0,01109	0,52
73	0,01190	0,52
74	0,01444	0,52
75	0,01584	0,3369
76	0,01800	0,3369
77	0,02086	0,3369
78	0,02314	0,3369
79	0,02752	0,3369
80	0,03171	0,3369
81	0,03773	0,3369
82	0,04558	0,3369
83	0,05041	0,3369
84	0,06084	0,3369
85	0,06900	0,214
86	0,08044	0,214
87	0,09341	0,214
88	0,11010	0,214
89	0,12545	0,214
90	0,14892	0,214

Fuente: Elaboración propia

Como se puede observar en la tabla 2.2, las tasas de consumo de alcohol en los últimos doce meses alcanzan su valor máximo en el período comprendido entre los 25 y los 34 años, por lo que el grupo de población que más ha consumido alcohol en el último año ha sido ese. A partir de los 35 años, las tasas se van reduciendo de forma gradual hasta llegar a los 90 años, que es la edad límite del estudio. Sorprendentemente, las mujeres más longevas siguen consumiendo alcohol, representando estas aproximadamente el 21% de la población femenina. Ahora bien, mientras que esta tasa se va reduciendo, las tasas de mortalidad, como es lógico, se incrementan a la vez que aumenta la edad. Por esta razón, el coeficiente de correlación lineal entre la mortalidad del año 2020 para mujeres y las tasas de consumo de alcohol en los últimos doce meses es negativo.

Figura 2.12

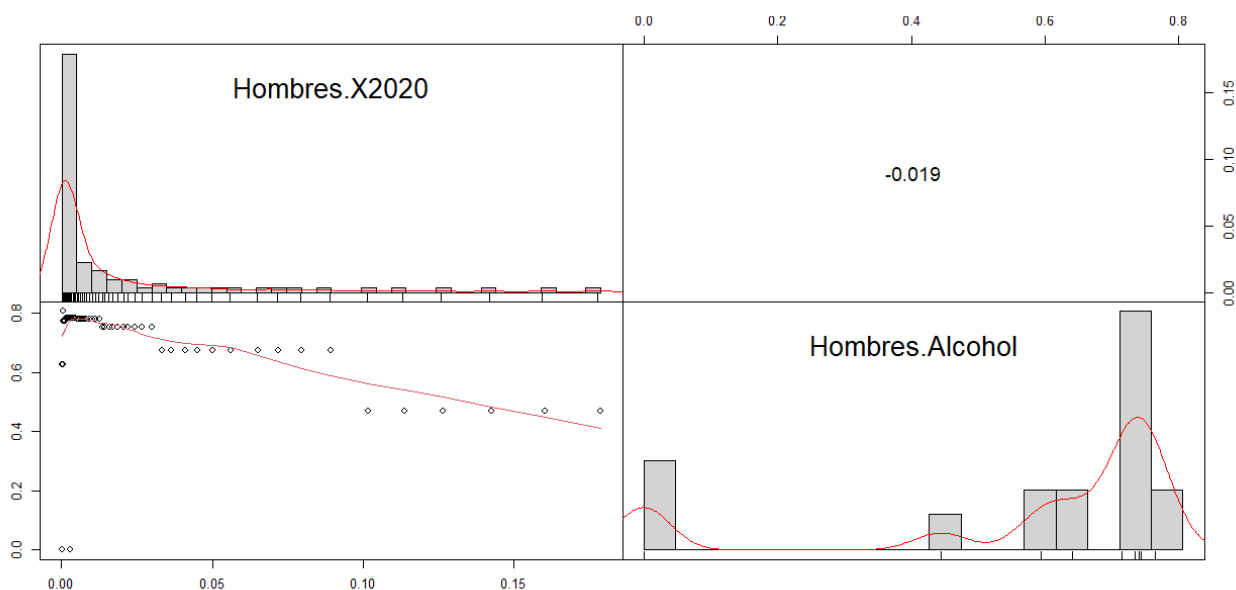
Matriz de correlaciones entre mortalidad masculina y consumo de alcohol en los últimos doce meses



Fuente: Elaboración propia

Figura 2.13

Matriz de correlaciones y significatividad entre mortalidad masculina y consumo de alcohol en los últimos doce meses



Fuente: Elaboración propia

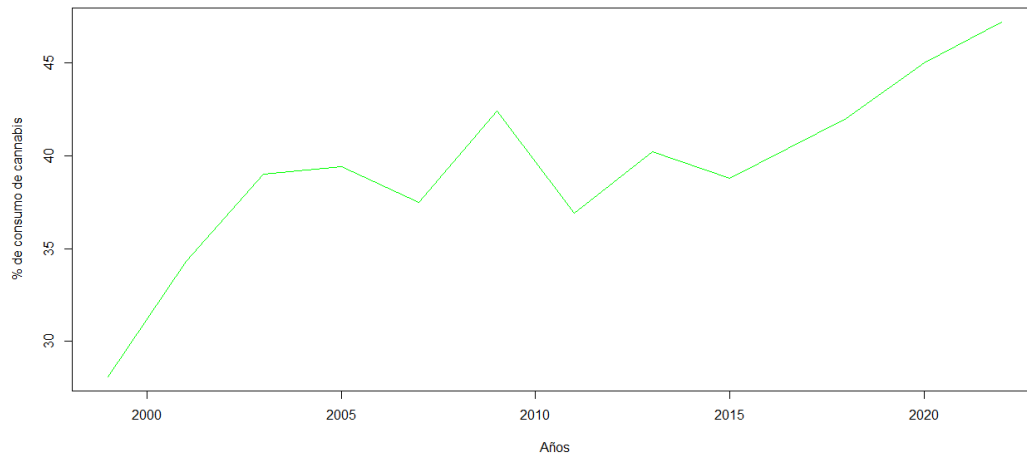
Las figuras 2.12 y 2.13 muestran la correlación entre la mortalidad de los hombres en el año 2020 y las tasas de consumo de alcohol en los últimos doce meses. En este caso, el efecto sería negativo y muy poco relevante. Sin embargo, no existen evidencias que demuestren que este factor tenga efectos significativos sobre la mortalidad del período.

El tercer factor es el consumo de drogas. Afecta de manera negativa a la salud. Las drogas no aumentan de forma instantánea las tasas de mortalidad, más bien muestran sus efectos en el largo plazo. Según el Plan Nacional sobre Drogas, algunos de estos efectos son: alteraciones cardiovasculares y neurológicas, psicosis, enfermedades broncopulmonares, algunos tipos de cáncer o esquizofrenia. Según el “Informe 2023 sobre alcohol, tabaco y drogas ilegales” del Observatorio Español de las Drogas y las Adicciones (OEDA), en el año 2021, se registraron más de mil muertes a consecuencia de una reacción grave a sustancias psicoactivas. El 78% de los fallecidos fueron hombres y el resto mujeres, de edad superior a 46 años (Observatorio Español de las Drogas y las Adicciones, 2023).

A continuación, se van a mostrar los gráficos, de elaboración propia, del consumo de las tres sustancias más relevantes en España según el Observatorio Español de las Drogas y las Adicciones.

Figura 2.14

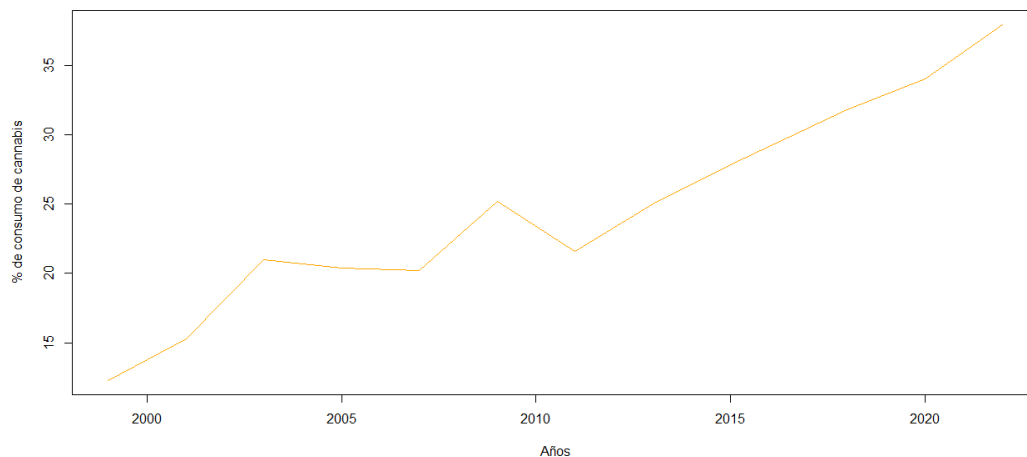
Consumo de cannabis en los últimos doce meses para población de 15 a 34 años, 1999-2022



Fuente: Elaboración propia

Figura 2.15

Consumo de cannabis en los últimos doce meses para población de 35 a 64 años, 1999-2022



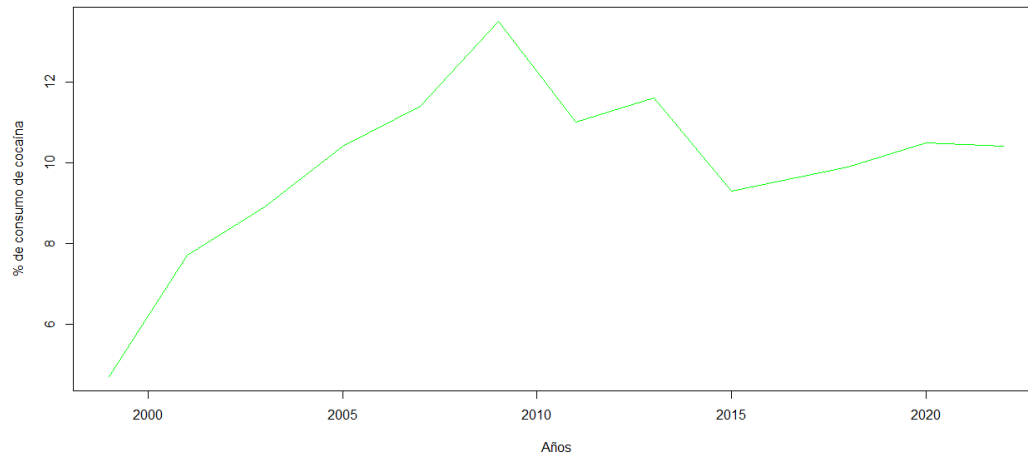
Fuente: Elaboración propia

Las figuras 2.14 y 2.15 muestran la evolución, entre 1999 y 2022, del porcentaje de población que ha consumido cannabis en los últimos 12 meses para dos grupos de edad, el primero entre los 15 y los 34 años y el segundo, entre los 35 y los 64. El patrón de consumo es similar en cuanto a los años, es decir, se da un crecimiento durante los primeros cinco años de la serie, después, sufre un ligero descenso, vuelve a aumentar hasta el año 2009, sufre otra caída y, finalmente, crece rápidamente hasta los últimos años con datos disponibles. Si se centra la atención en el año 2020, cabe recalcar que el

consumo es mayor entre los 15 y los 34 años (45% de la población) que entre los 35 y los 64 (34% de la población).

Figura 2.16

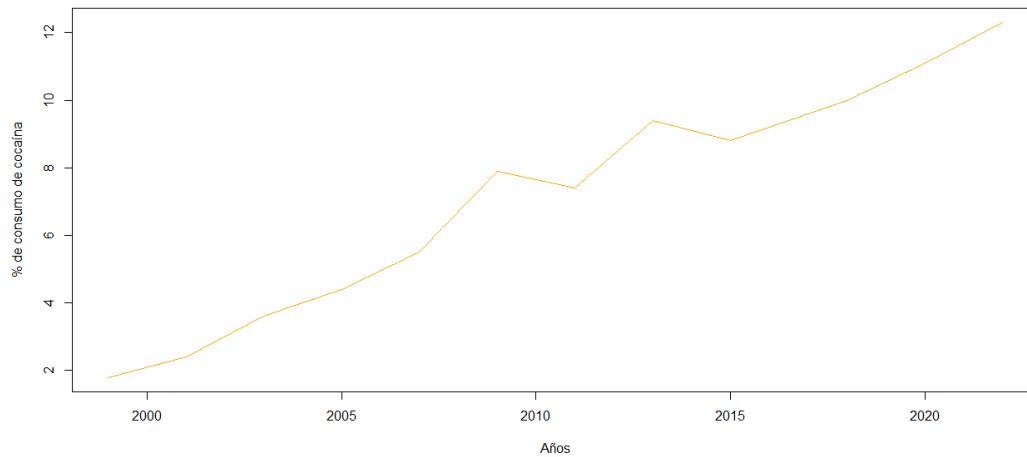
Consumo de cocaína en los últimos doce meses para población de 15 a 34 años, 1999-2022



Fuente: Elaboración propia

Figura 2.17

Consumo de cocaína en los últimos doce meses para población de 35 a 64 años, 1999-2022



Fuente: Elaboración propia

Las figuras 2.16 y 2.17 muestran la evolución, entre 1999 y 2022, del porcentaje de población que ha consumido cocaína en los últimos 12 meses para dos grupos de edad, el primero entre los 15 y los 34 años y el segundo, entre los 35 y los 64. En este caso, el patrón de consumo es muy diferente entre los grupos de edad. Para el primer grupo de

edad (es decir, entre los 15 y los 34 años) se da un máximo en el año 2009, punto a partir del cual, se va reduciendo el consumo con algún repunte, y se mantiene a partir de 2020. En cuanto al segundo grupo de edad (35 a 64 años), el patrón de consumo recuerda al del cannabis para el mismo grupo, siendo mayormente creciente, con alguna caída en 2010 y 2012. Sin embargo, el porcentaje de población que consume esta sustancia es aproximadamente una tercera parte de quienes consumían la anterior. Focalizando en 2020, el primer grupo de edad refleja un consumo del 10% y el segundo, del 11%.

Figura 2.18

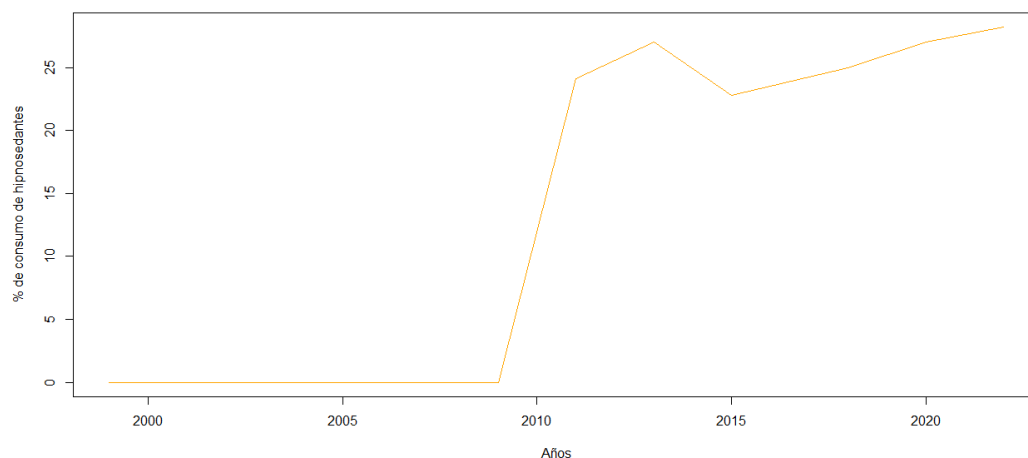
Consumo de hipnosedantes en los últimos doce meses para población de 15 a 34 años, 1999-2022



Fuente: Elaboración propia

Figura 2.19

Consumo de hipnosedantes en los últimos doce meses para población de 35 a 64 años, 1999-2022

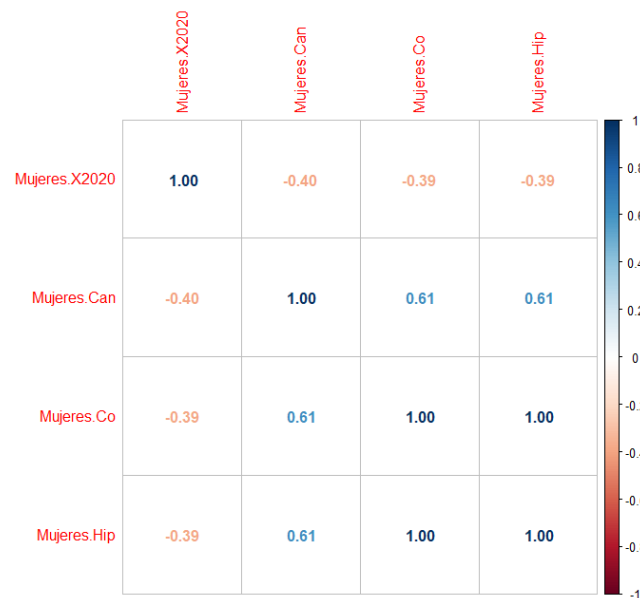


Fuente: Elaboración propia

Finalmente, las figuras 2.18 y 2.19 muestran el patrón de consumo entre 1999 y 2022 de hipnosedantes para los grupos de edad de 15 a 34 años y de 35 a 64 años. A diferencia de las anteriores, el consumo de esta sustancia no consta hasta el año 2009, año a partir del cual crece de manera muy significativa. El patrón de consumo es idéntico para ambos grupos de edad, si bien es cierto que el consumo del segundo grupo duplica al del primero. Las figuras 2.20 y 2.21 muestran la correlación lineal entre las tasas de mortalidad de las mujeres en el año 2020 y las tasas de consumo de las tres drogas más significativas en España.

Figura 2.20

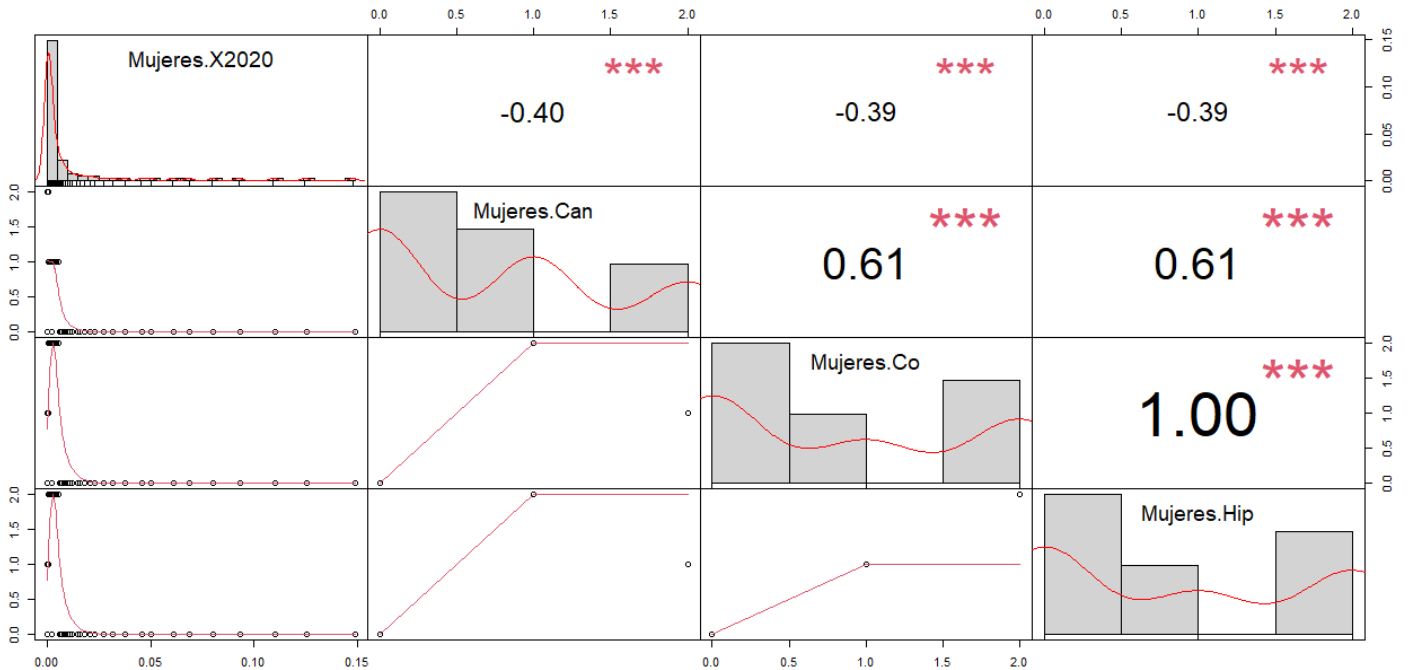
Matriz de correlaciones entre mortalidad femenina y consumo de drogas



Fuente: Elaboración propia

Figura 2.21

Matriz de correlaciones y significatividad entre mortalidad femenina y consumo de drogas



Fuente: Elaboración propia

Al igual que ha ocurrido con las tasas de tabaquismo y alcoholismo anteriormente mencionadas, las tasas de consumo de drogas parecen tener una correlación contraria a la esperada. A continuación, se encuentra la tabla 2.3, con información acerca de estas tasas y su relación con la mortalidad y la edad.

Tabla 2.3
Tasas de consumo de drogas y mortalidad

Edad	Mortalidad_2020	Can	Co	Hip
0	0,00228	0	0	0
1	0,00017	0	0	0
2	0,00008	0	0	0
3	0,00009	0	0	0
4	0,00002	0	0	0
5	0,00005	0	0	0
6	0,00006	0	0	0
7	0,00008	0	0	0
8	0,00008	0	0	0
9	0,00004	0	0	0
10	0,00005	0	0	0
11	0,00007	0	0	0
12	0,00006	0	0	0
13	0,00005	0	0	0
14	0,00010	0	0	0
15	0,00011	0,45	0,105	0,131
16	0,00011	0,45	0,105	0,131
17	0,00013	0,45	0,105	0,131
18	0,00012	0,45	0,105	0,131
19	0,00011	0,45	0,105	0,131
20	0,00012	0,45	0,105	0,131
21	0,00011	0,45	0,105	0,131
22	0,00012	0,45	0,105	0,131
23	0,00014	0,45	0,105	0,131
24	0,00016	0,45	0,105	0,131
25	0,00017	0,45	0,105	0,131
26	0,00019	0,45	0,105	0,131
27	0,00020	0,45	0,105	0,131
28	0,00020	0,45	0,105	0,131
29	0,00019	0,45	0,105	0,131
30	0,00022	0,45	0,105	0,131
31	0,00026	0,45	0,105	0,131
32	0,00029	0,45	0,105	0,131
33	0,00029	0,45	0,105	0,131
34	0,00026	0,45	0,105	0,131
35	0,00033	0,34	0,111	0,27
36	0,00039	0,34	0,111	0,27
37	0,00045	0,34	0,111	0,27
38	0,00042	0,34	0,111	0,27
39	0,00044	0,34	0,111	0,27
40	0,00059	0,34	0,111	0,27
41	0,00052	0,34	0,111	0,27
42	0,00067	0,34	0,111	0,27
43	0,00071	0,34	0,111	0,27
44	0,00077	0,34	0,111	0,27
45	0,00090	0,34	0,111	0,27
46	0,00095	0,34	0,111	0,27
47	0,00107	0,34	0,111	0,27
48	0,00119	0,34	0,111	0,27
49	0,00139	0,34	0,111	0,27
50	0,00151	0,34	0,111	0,27
51	0,00158	0,34	0,111	0,27
52	0,00200	0,34	0,111	0,27
53	0,00207	0,34	0,111	0,27
54	0,00225	0,34	0,111	0,27
55	0,00249	0,34	0,111	0,27
56	0,00277	0,34	0,111	0,27
57	0,00313	0,34	0,111	0,27
58	0,00345	0,34	0,111	0,27
59	0,00355	0,34	0,111	0,27
60	0,00405	0,34	0,111	0,27
61	0,00415	0,34	0,111	0,27
62	0,00453	0,34	0,111	0,27
63	0,00515	0,34	0,111	0,27
64	0,00546	0,34	0,111	0,27
65	0,00581	0	0	0
66	0,00633	0	0	0
67	0,00699	0	0	0
68	0,00729	0	0	0
69	0,00788	0	0	0
70	0,00888	0	0	0
71	0,00962	0	0	0
72	0,01109	0	0	0
73	0,01190	0	0	0
74	0,01444	0	0	0
75	0,01584	0	0	0
76	0,01800	0	0	0
77	0,02086	0	0	0
78	0,02314	0	0	0
79	0,02752	0	0	0
80	0,03171	0	0	0
81	0,03773	0	0	0
82	0,04558	0	0	0
83	0,05041	0	0	0
84	0,06084	0	0	0
85	0,06900	0	0	0
86	0,08044	0	0	0
87	0,09341	0	0	0
88	0,11010	0	0	0
89	0,12545	0	0	0
90	0,14892	0	0	0

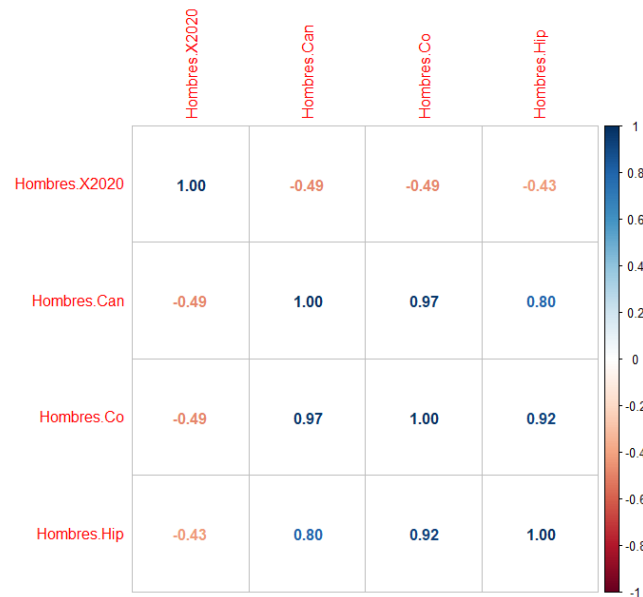
Fuente: Elaboración propia

En este caso, la problemática no se encuentra en que las tasas vayan disminuyendo con el paso de los años (aunque en el caso del cannabis sí podría ser una causa), si no que el problema reside en la falta de datos, pues únicamente se tiene información de la población que consume estas drogas entre los 15 y los 64 años. Es por esto que se correlacionan de forma negativa y moderada, ya que, según la información a partir de los 65 años, la población en general no consume drogas.

Además, otro punto a recalcar es que, con el paso del tiempo, el consumo de cocaína e hipnosedantes va aumentando, a diferencia de lo que ocurría con el tabaco, el alcohol, o el cannabis. Asimismo, como bien se comentaba en los efectos que tienen las drogas en la población, muchos de ellos son a largo plazo y este análisis no puede determinarlos, ya que utiliza la mortalidad del año 2020 para hacer una proyección. No obstante, este inconveniente podría solucionarse tomando como objeto de estudio las edades comprendidas entre los 18 y los 64 años, aunque esto provocaría que el análisis fuera incompleto.

Figura 2.22

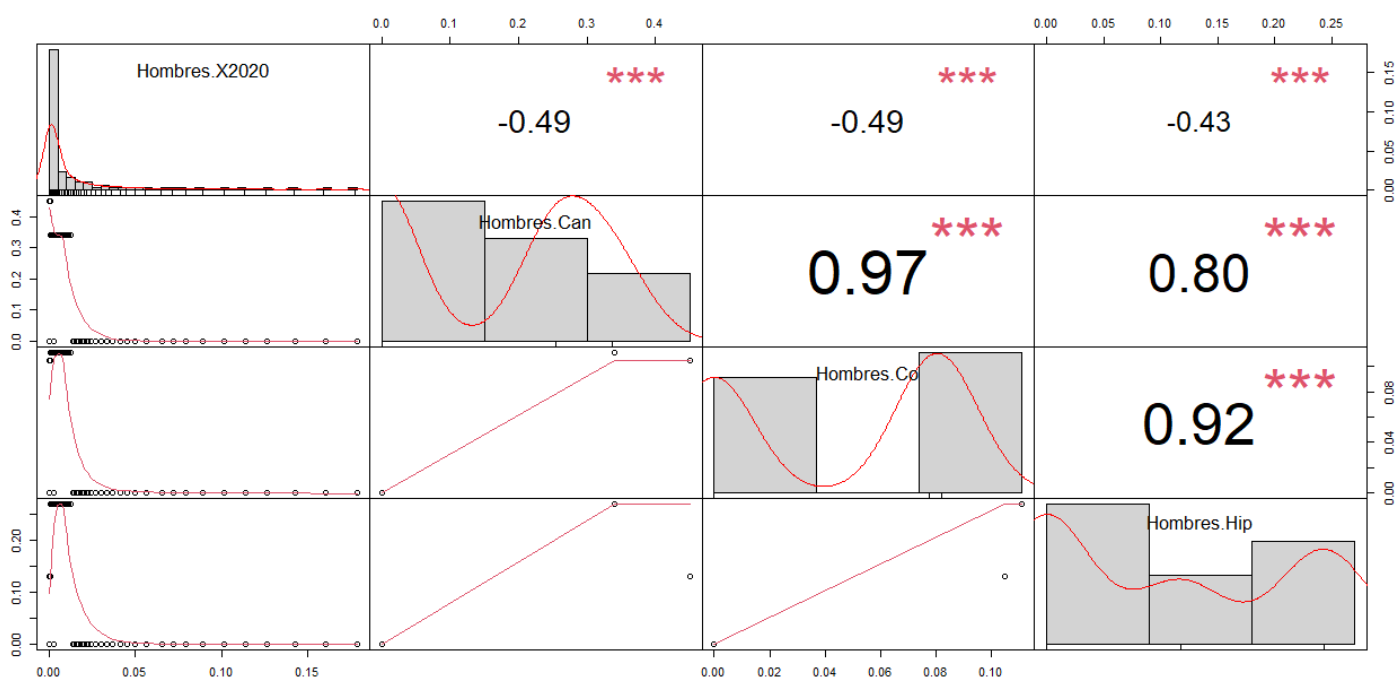
Matriz de correlaciones entre mortalidad masculina y consumo de drogas



Fuente: Elaboración propia

Figura 2.23

Matriz de correlaciones y significatividad entre mortalidad masculina y consumo de drogas



Fuente: Elaboración propia

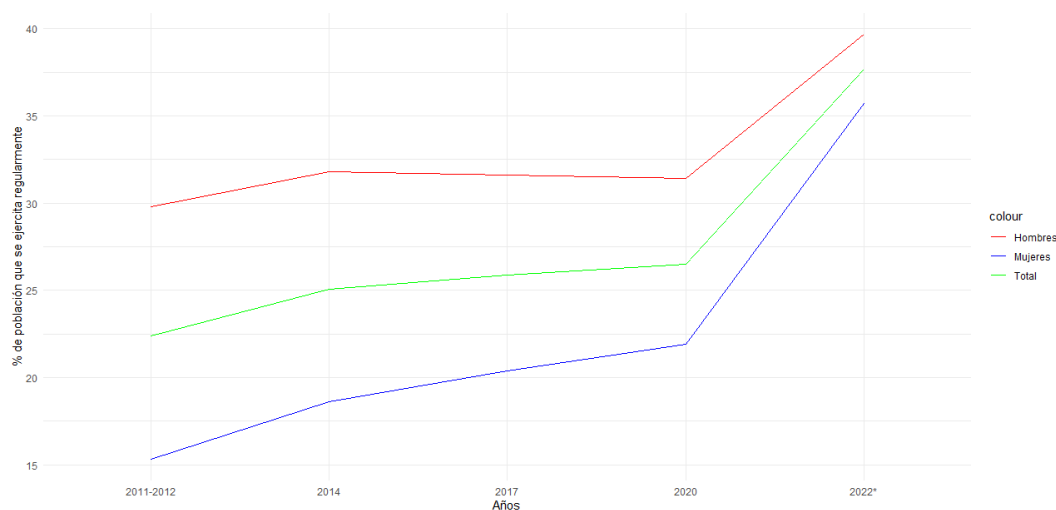
Aquí se muestran las correlaciones lineales entre la mortalidad de los hombres en el año 2020 y el consumo de las tres drogas más significativas del país. Para la población masculina, los efectos tienen las mismas direcciones que en el caso femenino, aunque sus coeficientes se acercan bastante más a -1.

El siguiente factor es la actividad física. Según la OMS, la realización de ejercicio físico habitual es una práctica muy beneficiosa para la salud, ya que reduce el riesgo de hipertensión, cardiopatías coronarias, diabetes, varios tipos de cáncer (como el de mama o el de colon) y mejora la salud ósea y el estado cardiorrespiratorio. Por ello, la inactividad física provoca el efecto contrario. Tanto es así, que las personas que no realizan el ejercicio físico suficiente están expuestas a un riesgo de mortalidad entre un 20% y un 30% mayor que aquellas personas que sí que realizan una cantidad de ejercicio físico recomendada (Organización Mundial de la Salud, 2019). Los datos recogidos por la Encuesta Nacional de Salud incluyen la actividad física que los entrevistados realizaron los siete días anteriores al estudio, es decir, los días por semana que se realiza ejercicio y el tiempo diario dedicado a practicarlo, siendo siempre superior a diez minutos diarios. Los resultados de la encuesta muestran un claro incremento en el porcentaje de población que realiza ejercicio físico regular, tanto para hombres como para mujeres, siendo en 2022

un 37,7% de hombres los que lo practican regularmente y un 35,7% de mujeres, mientras que en 2017 las tasas eran 31,6% y 20,4%, respectivamente.

Por su parte, la alimentación es un factor imprescindible para la salud. Sin embargo, dada la dificultad para cuantificar su efecto, se asumirá en este análisis que las personas que realizan ejercicio físico mantienen una alimentación acorde al número de veces que se ejercitan semanalmente. En caso de llevar a cabo una dieta poco saludable, se puede aumentar el riesgo de desarrollar enfermedades como diabetes, hipertensión, determinados tipos de cáncer, problemas cardiovasculares, etc. La OMS recalca que alrededor de 40 millones de niños menores de 5 años se encuentran en una situación de sobrepeso. Además, a causa del sobrepeso, cada año mueren más de 2,8 millones de personas (Organización Mundial de la Salud, 2024). Los resultados de la Encuesta Nacional de Salud en España, en 2017, reflejaron que un 44,3% de hombres y un 30% de mujeres padecen sobrepeso. A su vez, la encuesta afirma que “el 58% de hombres y el 69% de mujeres consumen fruta fresca a diario, y el 46% de mujeres y el 35% de hombres, verduras y hortalizas” (Ministerio de Sanidad, 2018).

Figura 2.24
Evolución de la práctica de ejercicio físico, 2011-2022



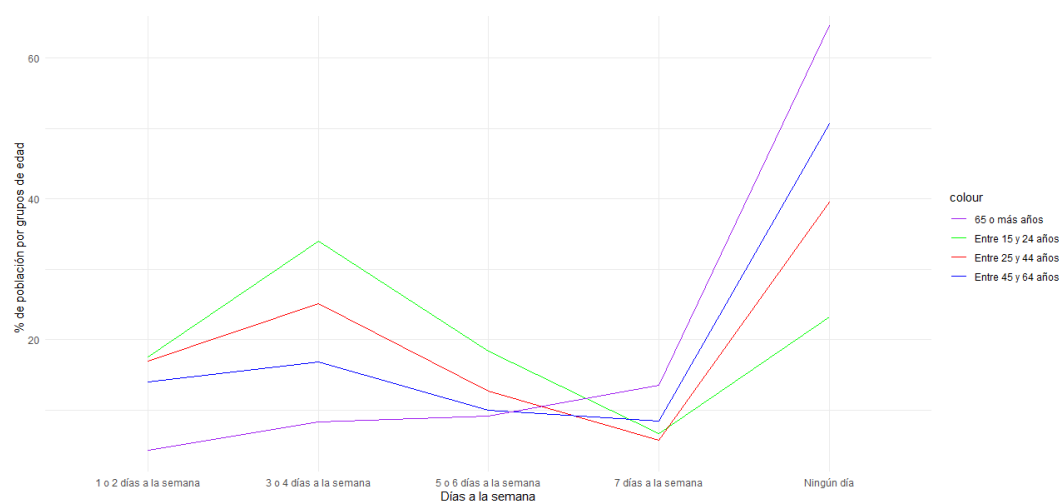
Fuente: Elaboración propia

La figura 2.24 muestra el porcentaje de población que realiza ejercicio físico regularmente, en una serie de tiempo que abarca desde el año 2011 hasta el 2022 distinguiendo entre hombres y mujeres. Los resultados reflejan que, durante los primeros

años de la serie, la población general llevaba a cabo poca actividad física regular, sobre todo las mujeres. No obstante, en el año 2020, se vivió un incremento repentino de las tasas de actividad física regular, tanto para hombres como para mujeres. Cabe destacar el hecho de que las mujeres deportistas en el año 2011 representaban únicamente al 11% de la población femenina, con una diferencia de un 15% con la masculina, pero para el año 2022, la diferencia se ha reducido a tan solo en 5% a pesar de que los hombres no han dejado de ejercitarse de forma regular, por lo que, se puede concluir que el incremento en la actividad física regular femenina ha sido mayor que la masculina.

Figura 2.25

Número de días de ejercicio físico semanal para hombres, por grupos de edad

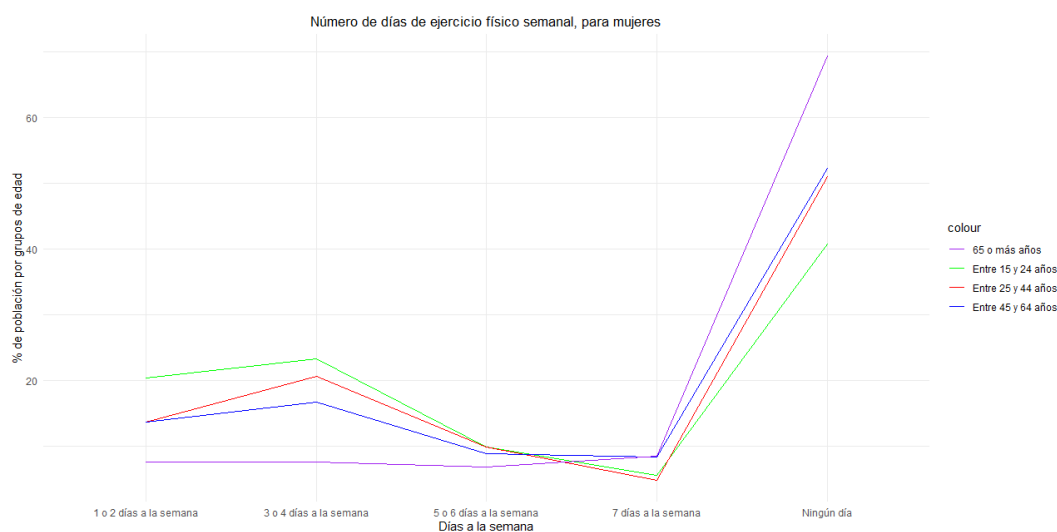


Fuente: Elaboración propia

La figura 2.25 muestra el número de días a la semana que los hombres dedican a realizar ejercicio, ordenados por grupos de edad. La mayoría de ellos no dedica ningún día a ejercitarse y muy pocos lo hacen 7 días semanales. De entre los que sí lo hacen, la mayor parte dedica entre 3 y 4 días a la semana, sobre todo aquellos con edades comprendidas entre los 15 y los 24 años. Como era de esperar, la población masculina mayor a 65 años es la que menos ejercicio realiza, en términos de porcentaje de población. Sin embargo, a pesar de esto, es el grupo de edad que más ejercicio realiza 7 días a la semana.

Figura 2.26

Número de días de ejercicio físico semanal para mujeres, por grupos de edad



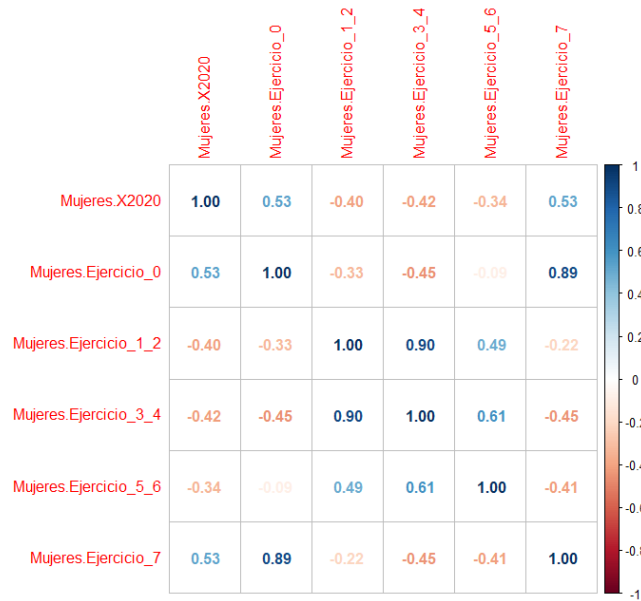
Fuente: Elaboración propia

Finalmente, la figura 2.26 muestra los días a la semana que las mujeres, por grupos de edad, realizan ejercicio físico. El patrón es muy similar al masculino, siendo el grupo de edad más joven el que dedica entre tres y cuatro días a la actividad física semanal. Una vez más, el grupo de mayor edad es el que dedica mayor tiempo semanal al ejercicio e, igual que pasaba con los hombres, la mayor parte de la población no dedica ningún día a entrenar.

Las figuras 2.27 y 2.28 muestran los coeficientes de correlación lineal entre la mortalidad de las mujeres en 2020 y la cantidad de días a la semana que practican ejercicio físico.

Figura 2.27

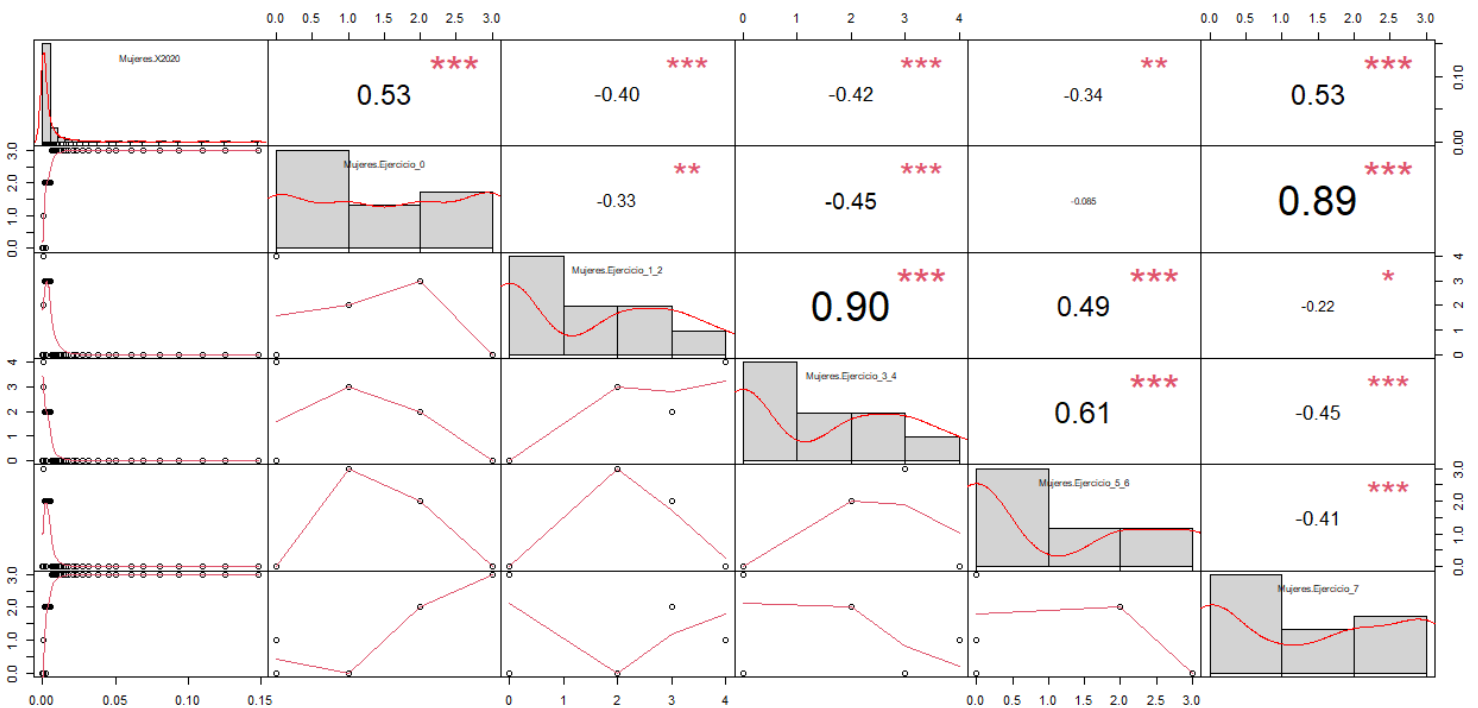
Matriz de correlaciones entre la mortalidad femenina y el ejercicio semanal



Fuente: Elaboración propia

Figura 2.28

Matriz de correlaciones y significatividad entre la mortalidad femenina y el ejercicio semanal



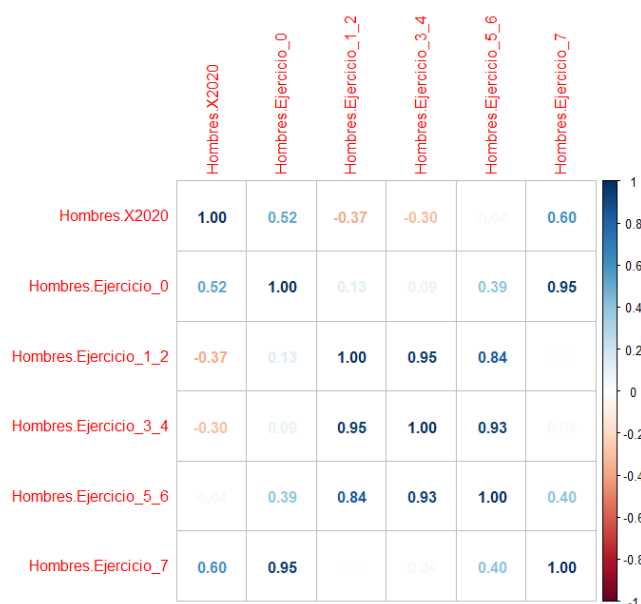
Fuente: Elaboración propia

Como era de esperar, la no realización de ejercicio físico tiene efectos muy negativos en la salud, concretamente, se ve que esta variable tiene una correlación de 0.53 con la mortalidad del período. Sorprendentemente, la realización de ejercicio físico durante todos los días de la semana también tiene efectos que favorecen la mortalidad femenina. Según estos datos, el tiempo de actividad física óptimo para reducir la mortalidad en mujeres sería de tres o cuatro días semanales, aunque la práctica de uno o dos días e incluso cinco o seis, es también beneficiosa para la salud. Como se comentaba en la descripción de esta variable, se asume que la alimentación de la población va en consonancia con la cantidad de actividad física que realiza.

Las figuras 2.29 y 2.30 muestran los coeficientes de correlación lineal entre la mortalidad masculina en el año 2020 y la cantidad de días semanales que practican ejercicio físico.

Figura 2.29

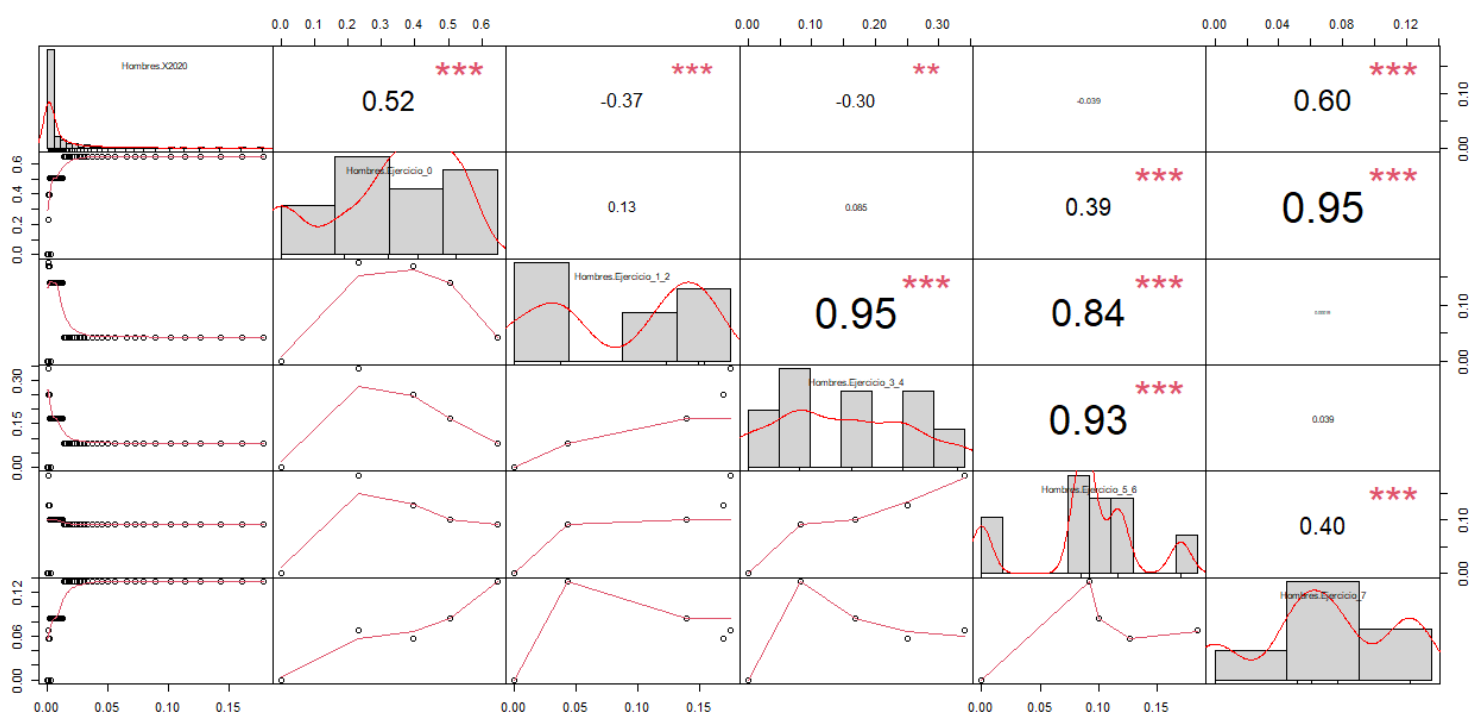
Matriz de correlaciones entre la mortalidad masculina y el ejercicio semanal



Fuente: Elaboración propia

Figura 2.30

Matriz de correlaciones y significatividad entre la mortalidad masculina y el ejercicio semanal



Fuente: Elaboración propia

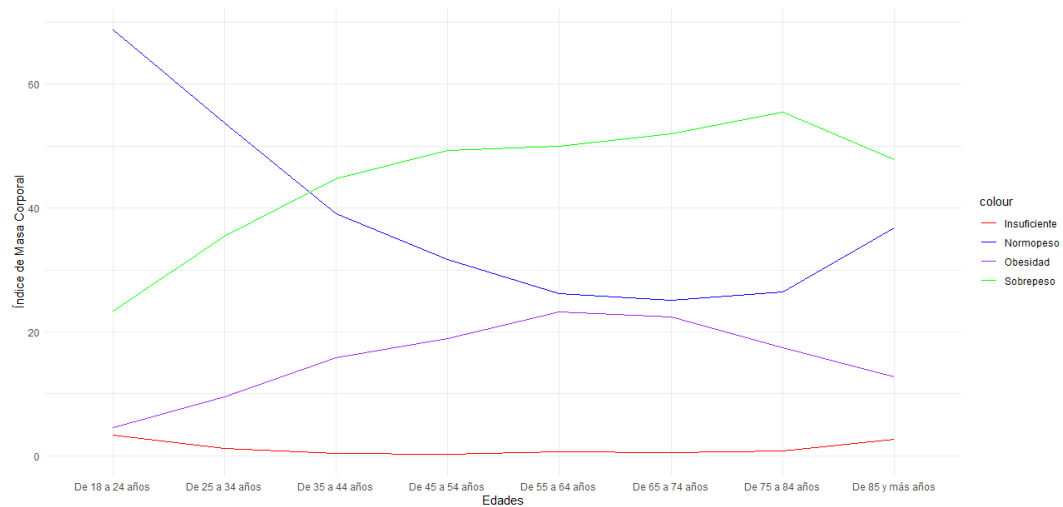
A diferencia de lo que sucedía con las mujeres, la práctica de cinco o seis días de ejercicio no resulta significativa bajo ningún nivel de confianza. El resto de las variables tiene efectos similares.

El siguiente factor analizado es el Índice de Masa Corporal (IMC). Es una medida que relaciona el peso de una persona con su altura, dando como resultado un porcentaje que indica los kilogramos de peso por cada metro al cuadrado de altura. De esta manera, se clasifican los resultados en peso insuficiente (cuando el IMC es inferior a 18,5), peso normal o normopeso (IMC situado entre 18,5 y 24,9), sobrepeso (cuando el IMC resulta entre 24,9 y 29,9) y obesidad (IMC superior a 30). Según la OMS, es recomendable mantener un Índice de Masa Corporal inferior a 25 y superior a 19, es decir, situarse en la medida de normopeso.

Este factor está relacionado con la actividad física y la alimentación y, como se mencionaba anteriormente, la OMS estimó en 2022 que “2.500 millones de adultos tenían sobrepeso, de los cuales 890 millones eran obesos. Esto significa que el 43% de los adultos (un 43% de hombres y un 44% de mujeres) tenían sobrepeso, lo que supone un aumento con respecto a 1990, cuando el porcentaje de adultos con sobrepeso era del 25%”

(Organización Mundial de la Salud, 2024). El sobrepeso y la obesidad surgen como respuesta a una alimentación poco saludable (con una ingesta de calorías alta) desequilibrada respecto a la actividad física que se realiza (gasto calórico bajo). La OMS documenta que, en 2019, un IMC alto provocó cinco millones de muertes por causas derivadas del sobrepeso y la obesidad, como enfermedades cardiovasculares, diabetes, cáncer y otras enfermedades respiratorias (Organización Mundial de la Salud, 2023).

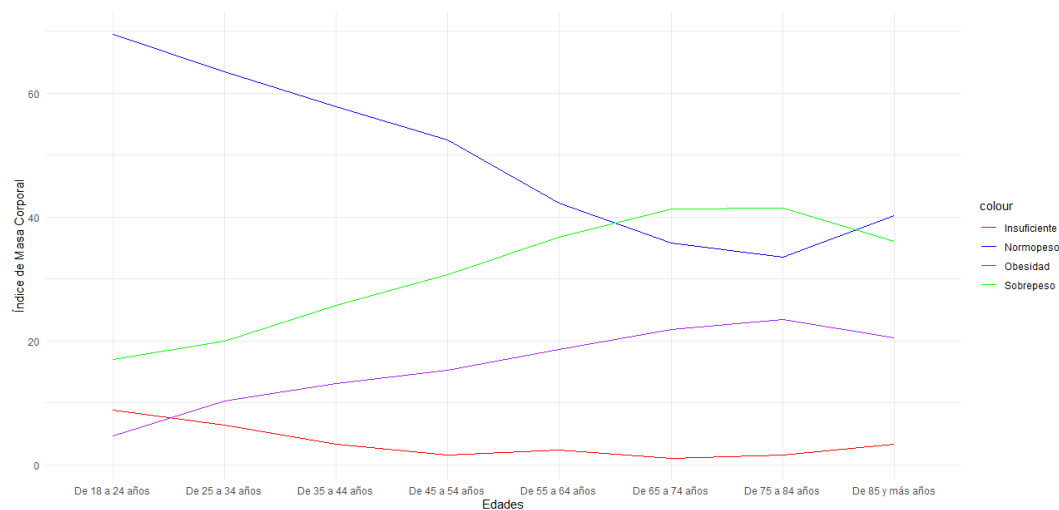
Figura 2.31
Índice de Masa Corporal por grupos de edad para hombres, 2020



Fuente: Elaboración propia

La figura 2.31 muestra el Índice de Masa Corporal de los hombres en 2020, agrupados por edades. En las edades más bajas domina el normopeso, pero, a medida que avanza la edad, el sobrepeso se convierte en el más común. Las tasas de población con un Índice de masa Corporal menor son muy bajas, incluso llegando a ser prácticamente 0 entre los 45 y los 54 años de edad, momento en el que la población sufre mayores tasas de obesidad. El pico del sobrepeso se encuentra entre los 75 y los 84 años, pero parece que existe una tendencia al normopeso cuando se supera ese umbral de edad.

Figura 2.32
Índice de Masa Corporal por grupos de edad para mujeres, 2020



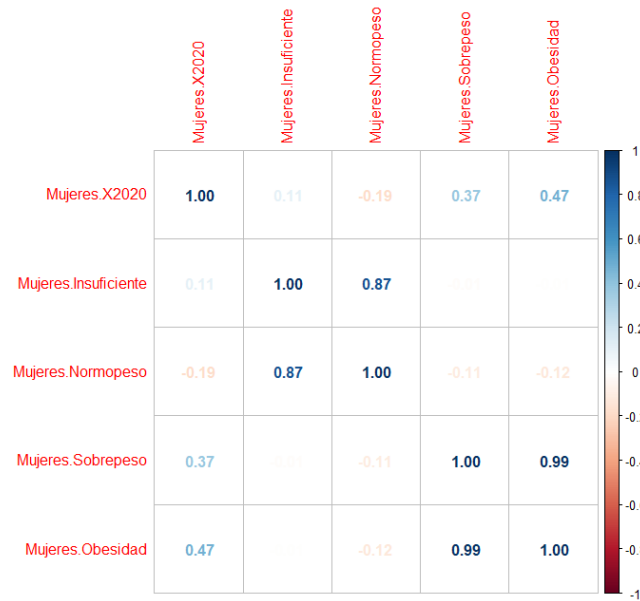
Fuente: Elaboración propia

Para terminar con el Índice de Masa Corporal, la figura 2.32 muestra el IMC por grupos de edad para las mujeres en 2020. Guarda ciertas similitudes con el gráfico de los hombres, pues la población menos representativa es la que tiene un índice más bajo. Sin embargo, las tasas de sobrepeso y normopeso convergen mucho más tarde, esta vez lo hacen entre los 55 y los 74 años y lo vuelven a hacer entre los 75 y los 85, mientras que para la población masculina lo hacían entre los 25 y los 44. Una vez más, se aprecia una ligera tendencia al normopeso a partir de los 85 años. Para finalizar, las tasas de obesidad son, en promedio, un 3% superiores en la población femenina.

Las figuras 2.33 y 2.34 muestran las correlaciones entre la mortalidad femenina de 2020 y las diferentes categorías de Índice de Masa Corporal.

Figura 2.33

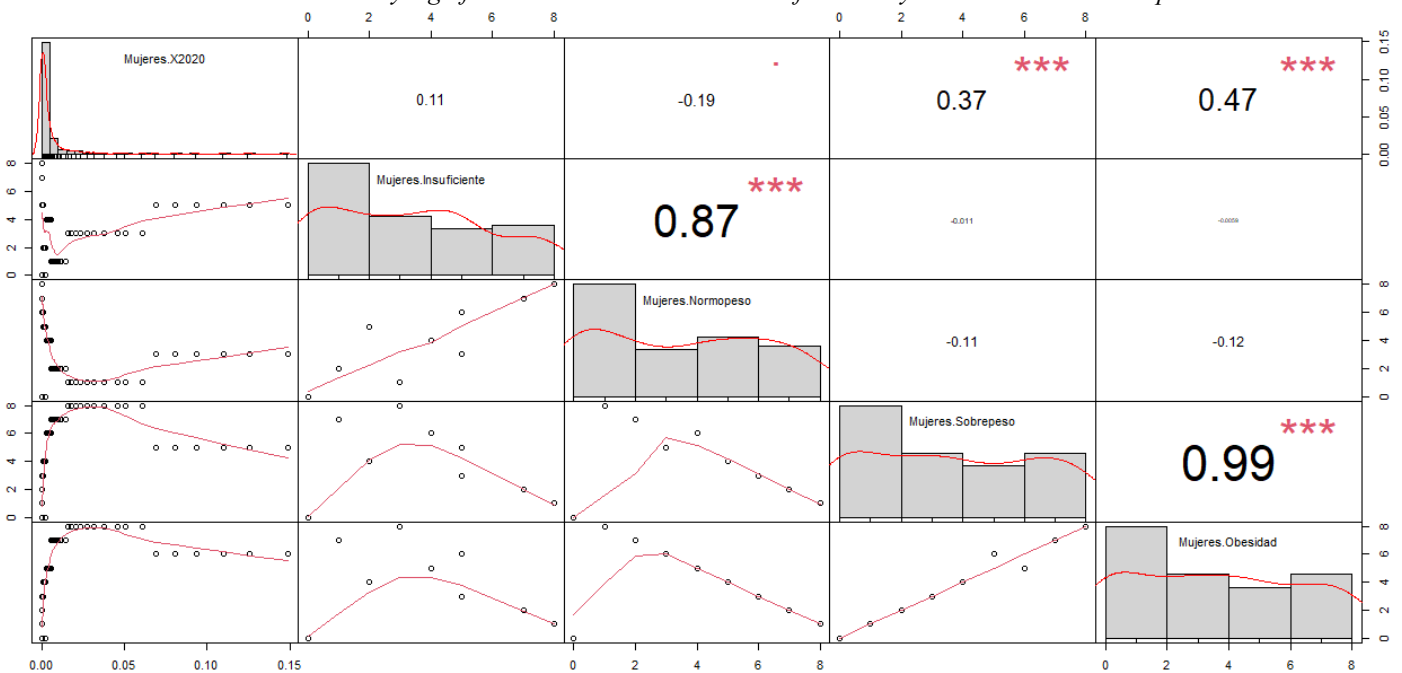
Matriz de correlaciones entre la mortalidad femenina y el Índice de Masa Corporal



Fuente: Elaboración propia

Figura 2.34

Matriz de correlaciones y significatividad entre la mortalidad femenina y el Índice de Masa Corporal



Fuente: Elaboración propia

Los efectos más significativos están relacionados con el sobrepeso y la obesidad, de manera que, sobre todo la obesidad, afectan de manera negativa en la salud, favoreciendo la mortalidad. En el caso del normopeso, tiene efectos contrarios al fallecimiento, aunque

a un nivel de confianza del 90%. Para el peso insuficiente no hay evidencias de que tenga efectos sobre la mortalidad.

Figura 2.35

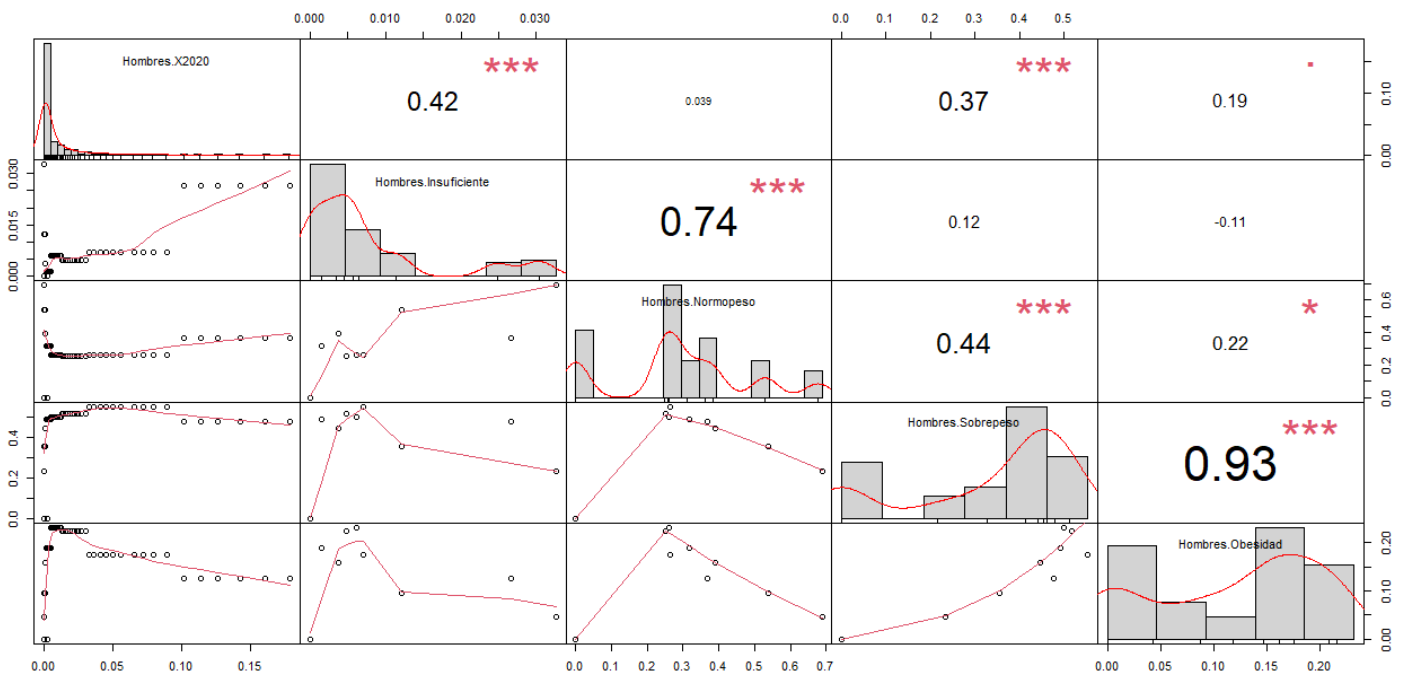
Matriz de correlaciones entre la mortalidad masculina y el Índice de Masa Corporal



Fuente: Elaboración propia

Figura 2.36

Matriz de correlaciones y significatividad entre la mortalidad masculina y el Índice de Masa Corporal



Fuente: Elaboración propia

En las figuras 2.35 y 2.36 se muestra la correlación entre la mortalidad masculina de 2020 y el índice de Masa Corporal. A diferencia de la población femenina, en este caso, la mayor significatividad la tienen el peso insuficiente y el sobrepeso, con efectos que favorecen la mortalidad. La obesidad también lo hace, aunque con un nivel de confianza del 90%. El normopeso parece no tener efectos significativos sobre las tasas de fallecimiento de los hombres.

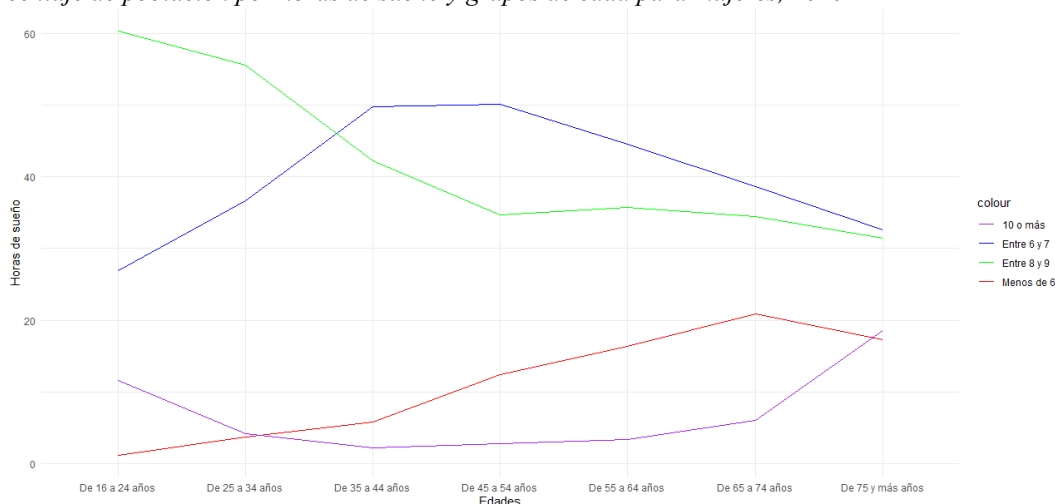
Otro factor son las horas de sueño. El Instituto Nacional del Corazón, los Pulmones y la Sangre de Estados Unidos (NHLBI son las siglas en inglés) estima que el tiempo recomendado de sueño diario para los adultos es de 7 u 8 horas (National Heart, Lung and Blood Institute, 2022).

Para cuantificar el sueño en este estudio, se han tratado las horas que duerme la población como una variable binaria, calculando para cada grupo de edad de los que se tienen datos el promedio de población que duerme más de 8 horas y el promedio de los que duermen menos de ese tiempo. Si el promedio de los que duermen más de 8 horas es mayor, la variable tendrá valor igual a 1, es decir, las horas que duerme ese grupo de edad, son suficientes y están en línea con las recomendaciones de los análisis sanitarios, mientras que, si dicho promedio es menor, la variable tomará valor igual a 0.

Los gráficos que se muestran a continuación son de elaboración propia y contienen los porcentajes de población divididos por sexo y grupos de edad, en función de las horas de sueño diario que mantienen.

Figura 2.37

Porcentaje de población por horas de sueño y grupos de edad para mujeres, 2020

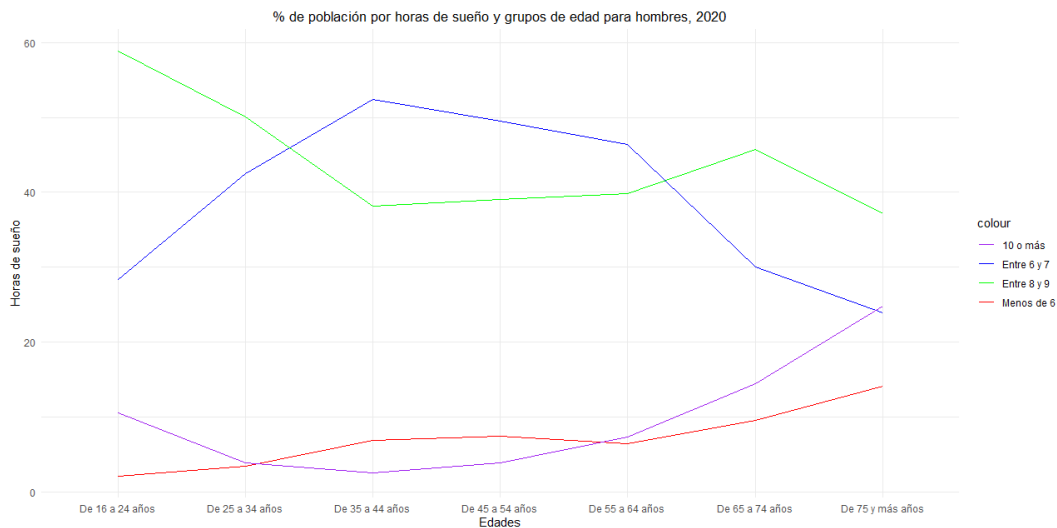


Fuente: Elaboración propia

La figura 2.37 contiene la información para las mujeres. Por lo general duermen entre 6 y 9 horas diarias, pero, como es lógico pensar, durante el periodo de actividad laboral es cuando mantienen un horario de sueño entre 6 y 7 horas.

Figura 2.38

Porcentaje de población por horas de sueño y grupos de edad para hombres, 2020



Fuente: Elaboración propia

Ahora se muestra el gráfico para los hombres. Lo más destacable a primera vista es que las tasas de población que duerme menos de 6 horas son mucho menores, sobre todo a partir de los 35 años. En promedio, los hombres duermen más que las mujeres, a excepción del tramo de edad entre los 65 y los 74 años, posiblemente a causa de cuestiones fisiológicas, como la menopausia y otros factores similares.

Las figuras 2.39 y 2.40 muestran la correlación entre la mortalidad de las mujeres en el año 2020 y las horas de sueño que duermen. Cabe recordar que esta variable ha sufrido una transformación y ahora es binaria, tomando el valor 1 cuando las horas de sueño son suficientes (es decir, más de 8 horas) y valor 0 en caso contrario.

Figura 2.39

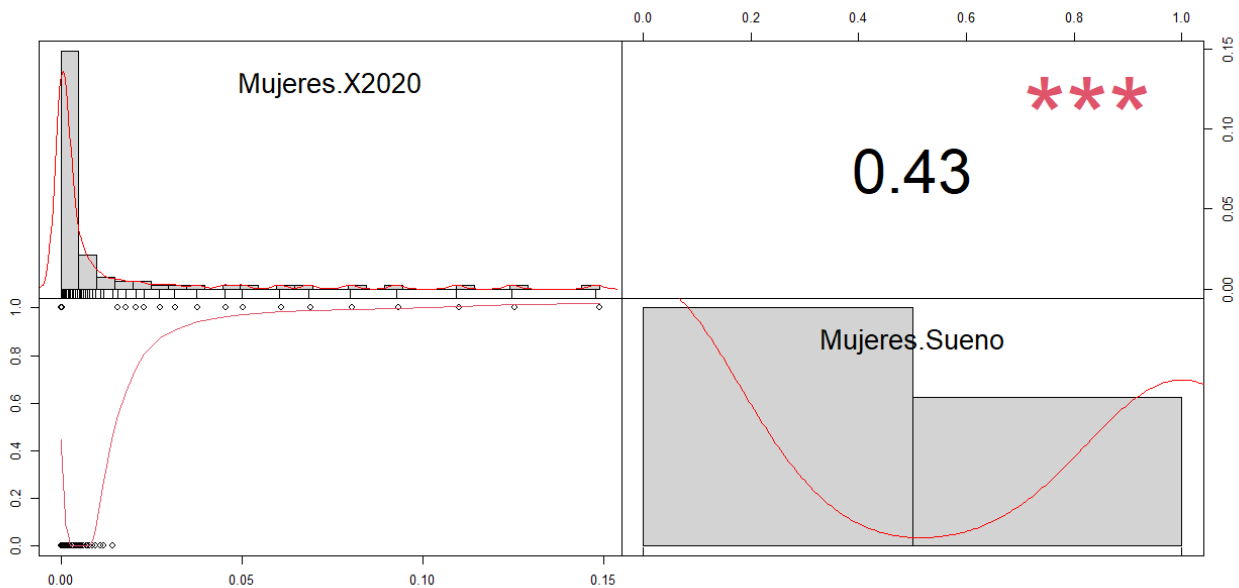
Matriz de correlaciones entre la mortalidad femenina y las horas de sueño



Fuente: Elaboración propia

Figura 2.40

Matriz de correlaciones y significatividad entre la mortalidad femenina y las horas de sueño



Fuente: Elaboración propia

El coeficiente de correlación resultante indica que las horas de sueño suficientes mantienen una relación moderada positiva con la mortalidad. Para entender mejor este efecto, se va a mostrar la tabla 2.4, que contiene las categorías de sueño y las tasas de mortalidad para cada edad.

Tabla 2.4*Tasas de mortalidad y edades que superan las horas de sueño recomendadas*

Edad	Mortalidad_2020	Sueno
0	0,00228	0
1	0,00017	0
2	0,00008	0
3	0,00009	0
4	0,00002	0
5	0,00005	0
6	0,00006	0
7	0,00008	0
8	0,00008	0
9	0,00004	0
10	0,00005	0
11	0,00007	0
12	0,00006	0
13	0,00005	0
14	0,00010	0
15	0,00011	0
16	0,00011	1
17	0,00013	1
18	0,00012	1
19	0,00011	1
20	0,00012	1
21	0,00011	1
22	0,00012	1
23	0,00014	1
24	0,00016	1
25	0,00017	1
26	0,00019	1
27	0,00020	1
28	0,00020	1
29	0,00019	1
30	0,00022	1
31	0,00026	1
32	0,00029	1
33	0,00029	1
34	0,00026	1
35	0,00033	0
36	0,00039	0
37	0,00045	0
38	0,00042	0
39	0,00044	0
40	0,00059	0
41	0,00052	0
42	0,00067	0
43	0,00071	0
44	0,00077	0
45	0,00090	0
46	0,00095	0
47	0,00107	0
48	0,00119	0
49	0,00139	0
50	0,00151	0
51	0,00158	0
52	0,00200	0
53	0,00207	0
54	0,00225	0
55	0,00249	0
56	0,00277	0
57	0,00313	0
58	0,00345	0
59	0,00355	0
60	0,00405	0
61	0,00415	0
62	0,00453	0
63	0,00515	0
64	0,00546	0
65	0,00581	0
66	0,00633	0
67	0,00699	0
68	0,00729	0
69	0,00788	0
70	0,00888	0
71	0,00962	0
72	0,01109	0
73	0,01190	0
74	0,01444	0
75	0,01584	1
76	0,01800	1
77	0,02086	1
78	0,02314	1
79	0,02752	1
80	0,03171	1
81	0,03773	1
82	0,04558	1
83	0,05041	1
84	0,06084	1
85	0,06900	1
86	0,08044	1
87	0,09341	1
88	0,11010	1
89	0,12545	1
90	0,14892	1

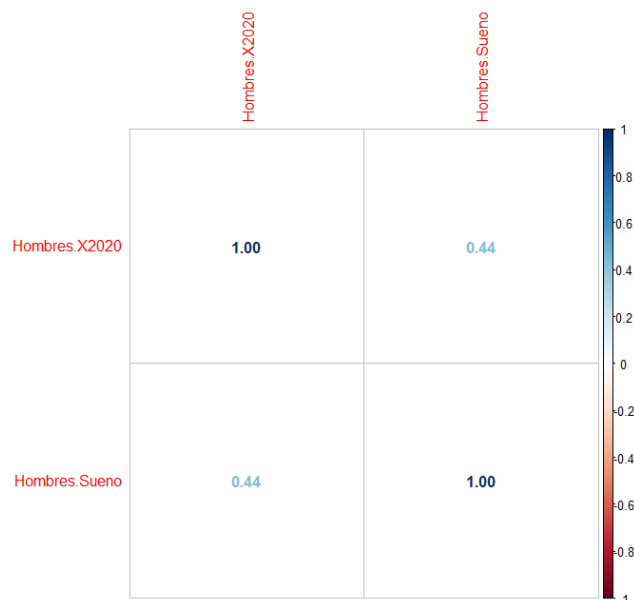
Fuente: Elaboración propia

Como se venía comentando, las horas de sueño son mayores en los tramos más avanzados de edad, pudiendo coincidir con las edades en las que la población femenina ya se encuentra jubilada, de forma que puede conciliar mejor el sueño. Los primeros registros no indican que las mujeres más jóvenes no cumplan las horas de sueño suficientes recomendadas por el NHLBI, si no que no se dispone de datos para la población menor de 15 años.

Las figuras 2.41 y 2.42 indican el coeficiente de correlación entre las tasas de mortalidad masculinas del año 2020 y las horas de sueño.

Figura 2.41

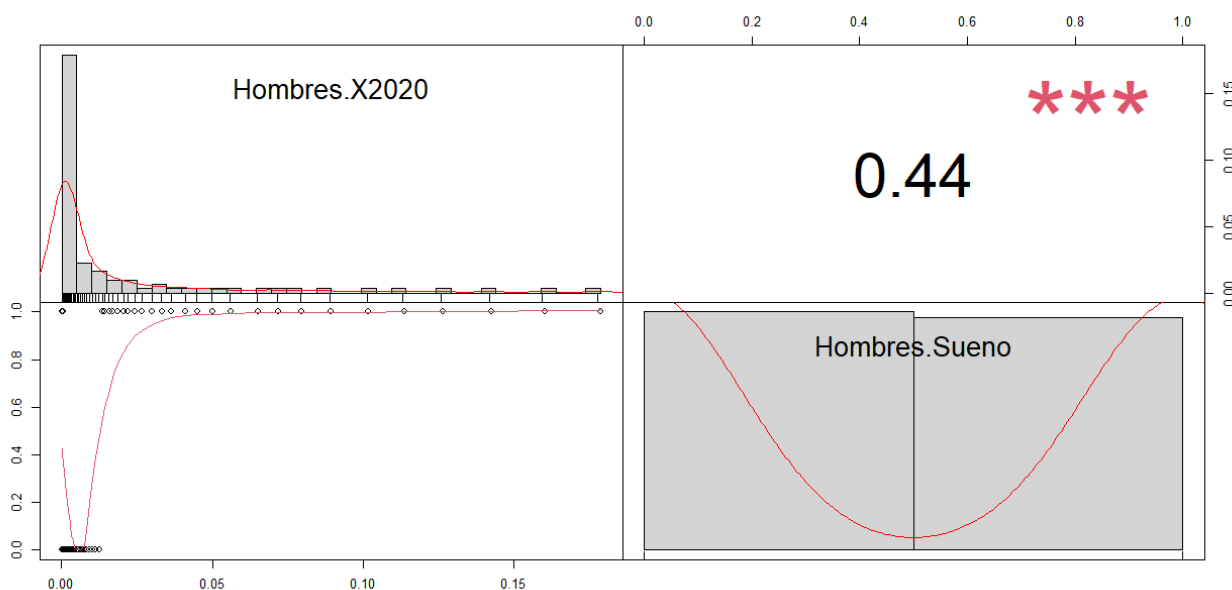
Matriz de correlaciones entre la mortalidad masculina y las horas de sueño



Fuente: Elaboración propia

Figura 2.42

Matriz de correlaciones y significatividad entre la mortalidad masculina y las horas de sueño



Fuente: Elaboración propia

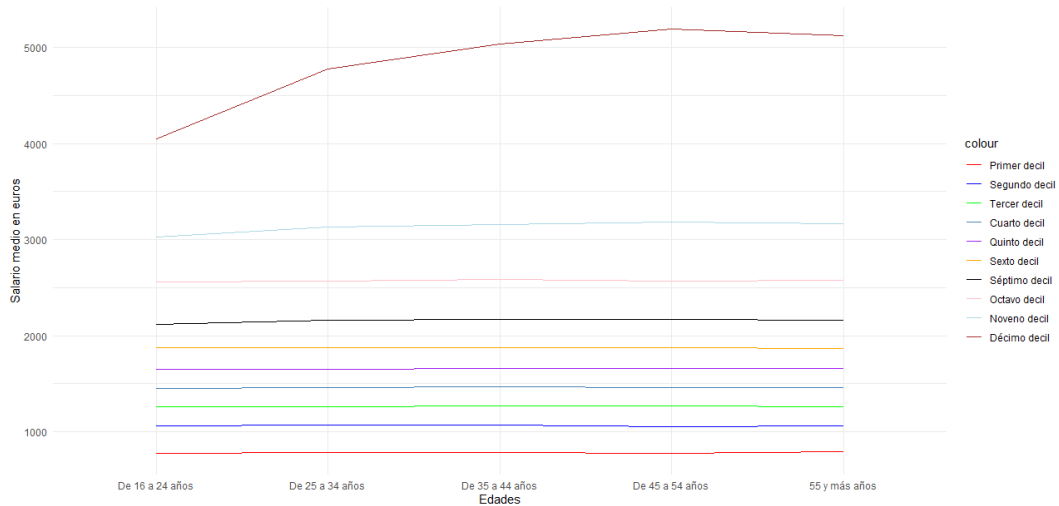
Los efectos son muy similares a los de la población femenina, porque la distribución de las horas de sueño sigue un patrón muy similar.

El siguiente factor es el nivel de renta. Puede influir de varias maneras en la mortalidad. En este caso, se estudiará el impacto de un nivel de renta bajo, es decir, aquel que se sitúa por debajo del percentil 70 del salario bruto mensual mediano, así como lo indica el INE (INE, 2023). Así, las personas con salarios más bajos suelen tener peor acceso a la atención médica de calidad, es decir, visitas regulares al médico, pruebas de detección preventiva o tratamientos médicos avanzados. De este modo, se podrían prevenir enfermedades y tratarlas de forma más efectiva, contribuyendo a tener una vida más longeva. Además, menores ingresos limitan a las personas para llevar a cabo hábitos saludables, como los comentados anteriormente en relación con la alimentación o la actividad física, así como a convivir en condiciones más desfavorables, como aire menos limpio, menos espacios verdes o peores infraestructuras.

Para este análisis se ha tratado el salario medio bruto mensual de España, asumiendo que no hay diferencias entre el percibido por mujeres y por hombres. Por otro lado, se ha utilizado la renta como una variable binaria, extrayendo el salario mediano de cada grupo de edad y comparándolo con el percentil 70 de la distribución de los salarios. Si para cada

grupo de edad el salario mediano excede dicho percentil, la variable renta tomará valor igual a 1, y si no lo hace, tendrá valor igual a 0.

Figura 2.43
Salarios medios brutos por grupos de edad y decil, 2021



Fuente: Elaboración propia

La figura 2.43 muestra los salarios medios brutos en España en 2021, por decil y grupos de edad. En general, los resultados son bastante lineales, es decir, se mantienen desde los 16 años hasta la jubilación, salvo los pertenecientes al noveno decil, que aumentan ligeramente a partir de los 25 años, y los del décimo decil, que crecen considerablemente entre los 16 y los 25 años, siguen creciendo de forma más lenta durante toda la vida del trabajador, hasta los 54 años, donde comienzan a disminuir de manera leve.

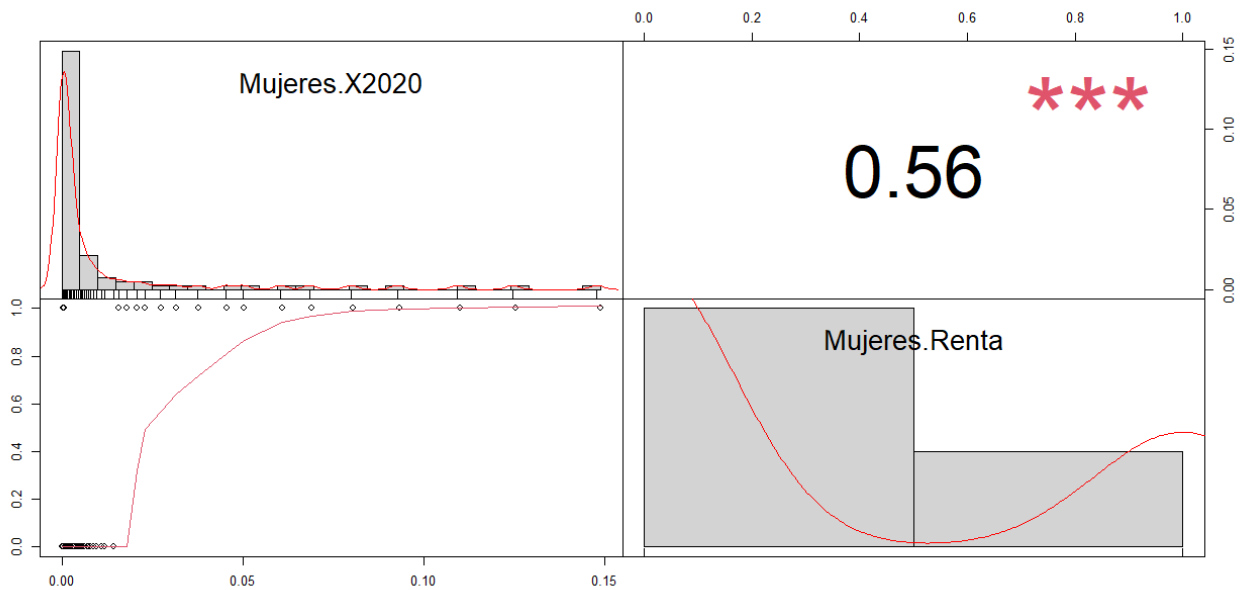
Las figuras 2.44 y 2.45 muestran el coeficiente de correlación lineal entre la mortalidad de las mujeres en 2020 y el nivel de renta.

Figura 2.44
 Matriz de correlaciones entre la mortalidad femenina y la renta



Fuente: Elaboración propia

Figura 2.45
 Matriz de correlaciones y significatividad entre la mortalidad femenina y la renta

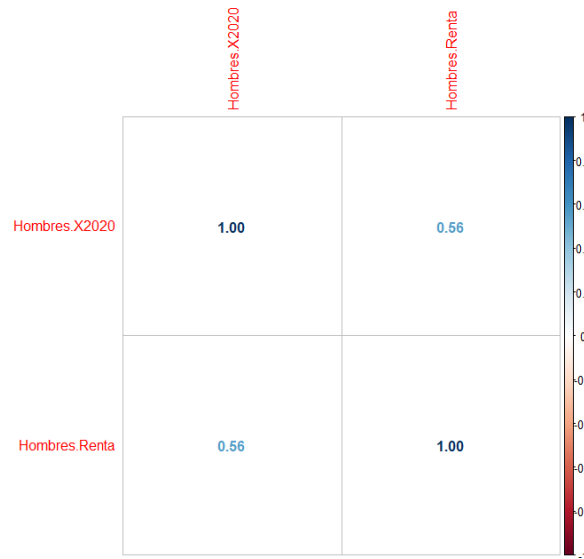


Fuente: Elaboración propia

A pesar de lo que se mencionaba en la descripción de la variable acerca del acceso a mejores tratamientos médicos con mayor renta, el coeficiente parece indicar que la mortalidad y los salarios guardan una relación positiva. Esto se debe a que, incluso asumiendo que se pueda acceder a mejor atención sanitaria, la mortalidad aumenta en

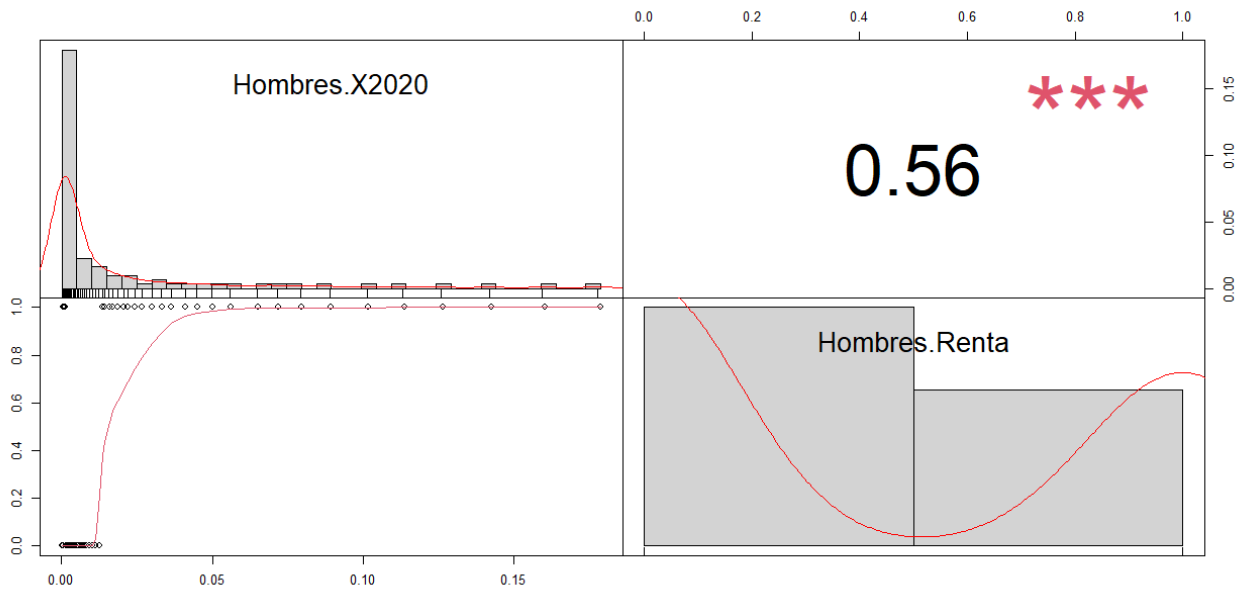
mayor medida cuando avanza la edad y, obviamente, no se puede explicar la supervivencia únicamente con un salario superior.

Figura 2.46
Matriz de correlaciones entre la mortalidad masculina y la renta



Fuente: Elaboración propia

Figura 2.47
Matriz de correlaciones y significatividad entre la mortalidad masculina y la renta



Fuente: Elaboración propia

Como se comentó en la descripción de la variable, se ha asumido que los salarios entre hombres y mujeres son los mismos. Por esta razón, los efectos de la renta sobre la mortalidad son idénticos para ambos sexos.

El último factor a tener en cuenta es la edad. De forma intuitiva, la edad es el primer factor de riesgo que viene a la mente a la hora de hablar de mortalidad. A medida que las personas envejecen, su riesgo de mortalidad aumenta por diversas razones. En primer lugar, el envejecimiento conlleva una mayor acumulación de enfermedades crónicas, como enfermedades cardíacas, diabetes y cáncer, que aumentan la probabilidad de muerte prematura. Además, el sistema inmunológico tiende a debilitarse con la edad, lo que hace que las personas mayores sean más susceptibles a infecciones graves y complicaciones. Los cambios fisiológicos y degenerativos asociados con el envejecimiento, como la pérdida de masa muscular y la disminución de la función de los órganos, también contribuyen al aumento del riesgo de mortalidad. La fragilidad física y la vulnerabilidad a lesiones, así como la menor capacidad de recuperación, son factores adicionales que hacen que las personas mayores sean más propensas a complicaciones graves que pueden resultar en muerte. En general, la edad influye en la mortalidad a través de una combinación de factores biológicos, físicos y sociales, lo que subraya la importancia de considerar la edad al evaluar el riesgo de mortalidad.

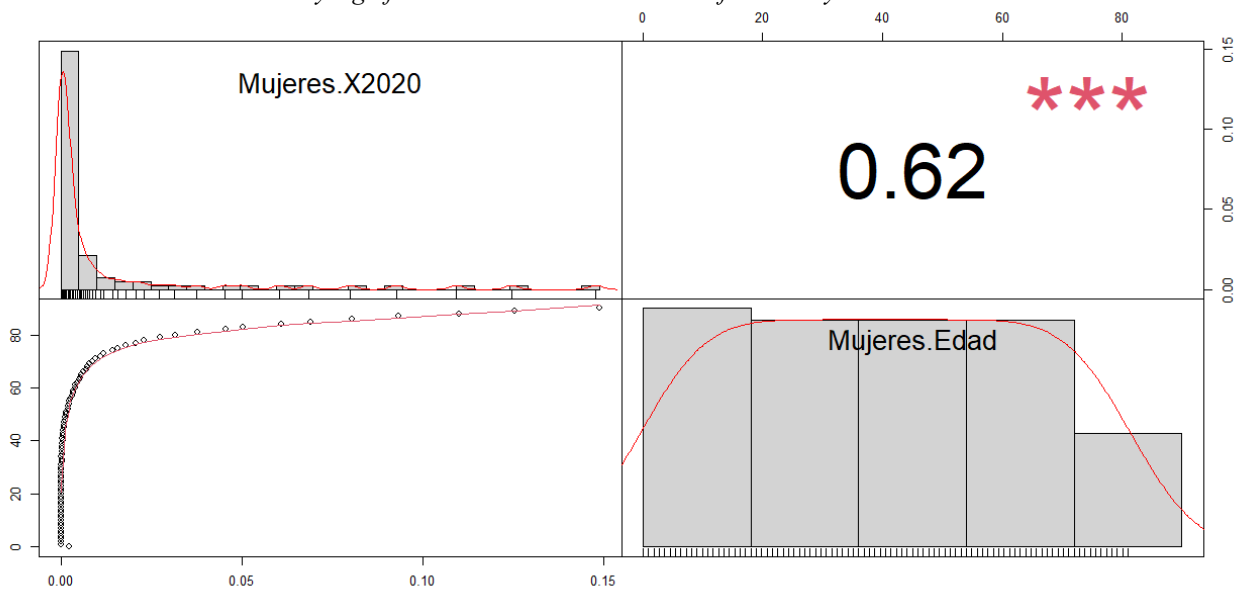
Las figuras 2.48 y 2.49 muestran el coeficiente de correlación entre la mortalidad femenina en el año 2020 y la edad.

Figura 2.48
Matriz de correlaciones entre la mortalidad femenina y la edad



Fuente: Elaboración propia

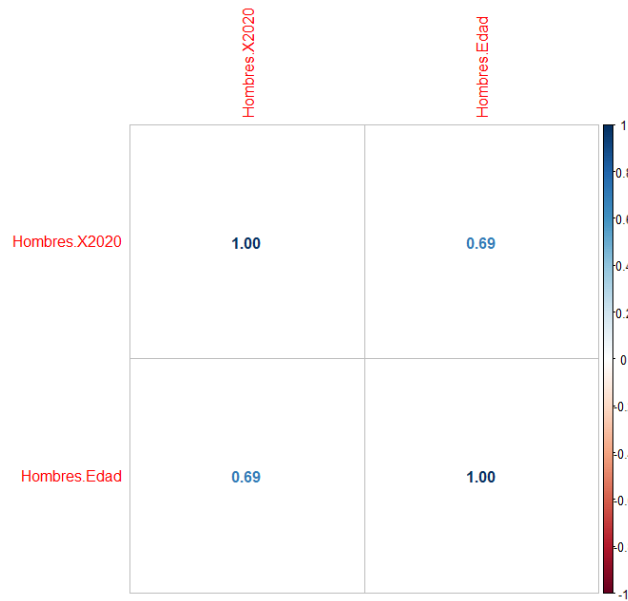
Figura 2.49
Matriz de correlaciones y significatividad entre la mortalidad femenina y la edad



Fuente: Elaboración propia

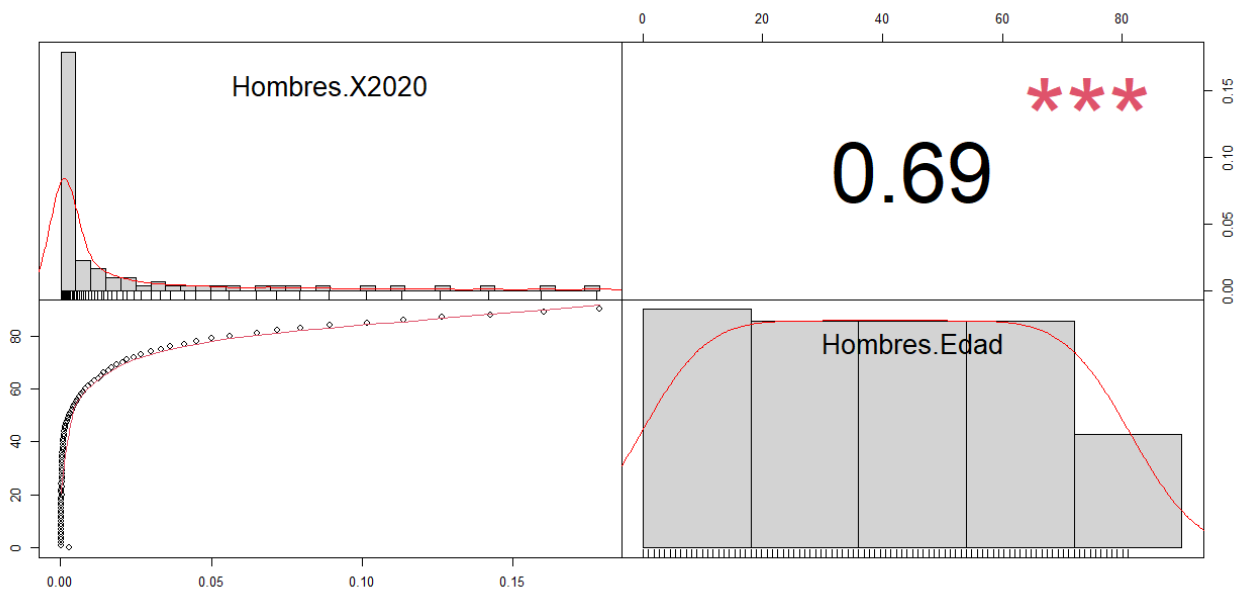
Este coeficiente es positivo, alto y significativo, por lo que, se podría afirmar que la mortalidad está estrechamente relacionada con la edad. Sin embargo, esta relación no es del 100%, es decir, que la mortalidad no está solo explicada por la edad, sino que también lo hace el resto de los factores de los que ya se ha habado.

Figura 2.50
Matriz de correlaciones entre la mortalidad masculina y la edad



Fuente: Elaboración propia

Figura 2.51
Matriz de correlaciones y significatividad entre la mortalidad masculina y la edad



Fuente: Elaboración propia

Las figuras 2.50 y 2.51 muestran la correlación lineal entre la mortalidad masculina en el año 2020 y la edad. Así como sucedía con las mujeres, se tiene un coeficiente de correlación alto, positivo y significativo, incluso mayor que el femenino. Esto indica que, para los hombres, la mortalidad está aún más relacionada con la edad que para las mujeres.

3. METODOLOGÍA

3.1. GLM

El Modelo Lineal Generalizado (GLM) es una extensión del modelo de regresión lineal simple, que permite modelar relaciones entre variables predictoras y una variable respuesta que, a diferencia de las regresiones lineales simples, no necesariamente tiene un término de error que siga una distribución normal. Este error puede ser binomial, gamma o Poisson. Además, el GLM permite que las varianzas no sean constantes y puede mantener una relación no lineal entre las variables predictoras y la respuesta, a través de una función de enlace. A continuación, se profundizará más en los conceptos y términos anteriores.

En cuanto a los errores, como se comentaba, no tienen por qué seguir una distribución normal. Antes de la existencia de los GLM, la falta de normalidad se trataba transformando la variable respuesta o utilizando métodos no paramétricos. Sin embargo, ahora se pueden emplear datos con errores con distintas distribuciones, como binomial (se tratan proporciones, como tasas de mortalidad, tasas de desempleo, porcentaje de éxito de un tratamiento, etc.), gamma (son datos cuya varianza aumenta a medida que lo hace la media de la muestra y de forma constante, como intervalos de tiempo entre un suceso y otro) o Poisson (son datos de conteo, como el número de fallecimientos o la población afectada por un factor de riesgo).

Para relacionar las variables predictoras con la variable respuesta, el GLM necesita tres componentes: una función de enlace, un predictor lineal y una distribución de probabilidad.

La función de enlace relaciona la media de la variable de respuesta con una combinación lineal de las variables predictoras. Tiene como propósito transformar la variable respuesta, para que pueda ser modelada por un modelo lineal. Por último, proporciona flexibilidad, ya que la relación entre las variables predictoras y la variable respuesta puede ser no lineal, siendo las funciones de enlace más usadas:

La identidad:

$$g(\mu) = \mu \quad (1)$$

El logaritmo:

$$g(\mu) = \log(\mu) \quad (2)$$

El logit:

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) \quad (3)$$

El predictor lineal incluye las variables predictoras y sus coeficientes, junto a la función de enlace y establece la relación existente entre ellas. Se representa a través de la ecuación:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (4)$$

Donde:

- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de regresión.
- X_1, X_2, \dots, X_n son las variables predictoras.

La distribución de probabilidad especifica la distribución de probabilidad de la variable respuesta condicionada por los valores predichos por el modelo lineal, y engloba cualquier distribución de probabilidad que se adecúe a los datos que se estén utilizando. Normalmente se utilizan las siguientes distribuciones: normal, binomial, Poisson, gamma o exponencial.

Para este análisis se utilizará el software R y se emplearán diferentes modelos GLM, que representarán a una persona con determinadas características, siempre diferentes al resto, para así contemplar distintas casuísticas futuras, con el objetivo de verificar si los factores de riesgo descritos en los apartados anteriores son significativos para estudiar la variación en la mortalidad. Para ello, se incluirán dichos factores como variables descriptivas, la mortalidad como variable respuesta y la distribución Gamma como función de enlace, ya que se tratará de calcular el impacto con las tasas de mortalidad en cada período. Las explicaciones más detalladas y los resultados se encuentran en la sección siguiente.

3.2. Inferencia y contraste de hipótesis en GLM

Ya se ha tratado la forma de los modelos GLM. Ahora, se explicará cómo se lleva a cabo la inferencia y el contraste de hipótesis de dichos modelos.

Los parámetros del modelo GLM se estiman utilizando métodos de optimización para maximizar la verosimilitud, que pueden basarse en la metodología de “Scoring” o en el Método de Newton-Raphson. Ambos son métodos de estimación iterativos y encuentran los valores de los parámetros que mejor ajustan el modelo a los datos observados. El

objetivo es hacer una estimación inicial de los parámetros del modelo e iterar hasta que converja en una solución. Cada una de las iteraciones implica el cálculo de los valores de los parámetros, empleando los valores anteriores y la primera y segunda derivada de la función de verosimilitud de los parámetros. Esta función de verosimilitud representa la probabilidad de obtener los datos observados dados los parámetros del modelo y su objetivo es encontrar los valores que maximicen la función, de manera que se encuentren aquellos valores que consiguen que los datos sean más probables bajo el modelo propuesto.

Una vez estimados los parámetros del modelo, es fundamental evaluar su significatividad. Para ello, se construyen intervalos de confianza para los parámetros, que proporcionan una estimación de la incertidumbre asociada a las estimaciones. Además, se realizan pruebas de hipótesis para determinar si los parámetros son significativamente diferentes de cero. Esto ayuda a comprender la contribución de cada variable independiente al modelo y su efecto sobre la variable dependiente. El contraste de hipótesis se utiliza para probar la importancia de los efectos de las variables independientes en el modelo. Se compara el ajuste del modelo completo, que incluye todas las variables predictoras, con el ajuste de un modelo reducido que excluye la variable de interés. La diferencia en la bondad de ajuste entre estos modelos se evalúa mediante una estadística de prueba, como la prueba de razón de verosimilitud o el estadístico de Wald. Si la estadística de prueba fuera significativa, se rechazaría la hipótesis nula de que el efecto de la variable es cero, lo que indicaría que la variable tendría un efecto significativo en el modelo.

Para seleccionar las variables relevantes y significativas, se pueden utilizar métodos de eliminación hacia atrás (backward elimination) o la selección hacia adelante (forward selection). Estos métodos se basan en criterios de información como el AIC o el BIC, que penalizan la complejidad del modelo para evitar el sobreajuste.

3.3. Modelo de Lee-Carter

El modelo de Lee-Carter es un método estadístico que se utiliza en el mundo actuarial para proyectar la mortalidad. Fue desarrollado en 1992 por Ronald Lee y Lawrence Carter, dos demógrafos estadounidenses. Este modelo se basa en la descomposición de la mortalidad en dos componentes principales: la tendencia temporal y una función de edad.

Tiene como objetivo capturar cómo cambia la mortalidad con el tiempo y cómo varía según la edad, y tiene la siguiente forma:

$$m_{x,t} = e^{(a_x + b_x k_t + \varepsilon_{x,t})} \quad (5)$$

De forma simplificada:

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \quad (6)$$

Donde:

- $m_{x,t}$ es la tasa de mortalidad correspondiente al año t y para cada edad x .
- a_x es un componente que depende únicamente de la edad x y no está influida por el tiempo t . Muestra el comportamiento de la curva de mortalidad para cada edad x .
- b_x es un coeficiente que captura cómo varía la mortalidad de cada edad x a medida que cambia el nivel de mortalidad general.
- k_t es un factor de tendencia temporal común a todas las edades, que explica cómo varía la mortalidad en el tiempo.
- $\varepsilon_{x,t}$ es el término de error para la edad x en el momento del tiempo t . Mide el riesgo de mortalidad para la situación recién descrita.

El propósito de este modelo es estimar los coeficientes a_x , b_x y k_t que se ajusten óptimamente a los datos de mortalidad disponibles. Una vez calculados, estos coeficientes se emplearán para realizar proyecciones futuras de la mortalidad.

La estimación de los coeficientes se lleva a cabo generalmente mediante técnicas de regresión, como la regresión lineal, utilizando los datos históricos de mortalidad. Una vez obtenidos los coeficientes, se aplican al modelo para proyectar la mortalidad futura, utilizando las tasas de mortalidad base esperadas y extrapolando la tendencia temporal.

Es importante señalar que el modelo de Lee-Carter se fundamenta en ciertas suposiciones cruciales, como la estacionalidad de la mortalidad y la estabilidad de las relaciones entre la tendencia temporal y la mortalidad específica por edad a lo largo del tiempo.

En este análisis se va a utilizar el software R para obtener las tasas de mortalidad estresadas por los factores de riesgo significativos obtenidos en los GLM anteriores. Se ha especificado una base de datos que incluye los períodos a los que está sujeto el estudio (1960 a 2020), las edades que contempla (11 a 64 años), la población expuesta a la mortalidad y las tasas de mortalidad para hombres y mujeres. El objetivo es predecir diez

años en el futuro cuáles van a ser las tasas de mortalidad si estas se encuentran afectadas por los factores de riesgo, es decir, para cada uno de los escenarios que representa cada GLM, se proyectará la mortalidad y se comparará con las tasas reales proyectadas, para obtener los efectos futuros de la influencia de los factores de riesgo en la mortalidad y comprobar si llevar a cabo una vida saludable es sinónimo de tener una vida más longeva y de una reducción de las tasas de mortalidad. Además, para verificar la consistencia del modelo predictor, se ha realizado una prueba de backtesting. Para ello, se ha extraído un train set, que representa las tasas de mortalidad para cada edad entre 1960 y 2010, y un test set, entre 2011 y 2020. Con esto, se aplicará el modelo Lee Carter para comprobar si las tasas de mortalidad del train set proyectadas en 2020 son las mismas que las tasas de mortalidad reales de ese mismo año.

3.4. Tablas de mortalidad

Las tablas de mortalidad proporcionan información detallada sobre la probabilidad de muerte para individuos de diferentes edades y géneros en una población determinada durante un período de tiempo específico. Se construyen a partir de datos recopilados de registros vitales, censos y encuestas demográficas e incluyen información sobre el número de nacimientos, defunciones y la población en riesgo en diferentes grupos de edad y género. A partir de estos datos se pueden calcular tasas de mortalidad específicas por edad y género, así como también otros indicadores demográficos importantes, como la esperanza de vida al nacer.

Las tablas de mortalidad incluyen una serie de funciones biométricas, que se emplean para el cálculo de las diferentes métricas que se expondrán a continuación. En concreto, dichas funciones son:

- Supervivencia a la edad x . Se representa como l_x y equivale al número de individuos que consiguen llegar con vida a la edad x .
- Fallecimiento con edad x . Se representa con d_x y conforma el número de fallecimientos para las personas que han llegado a cumplir la edad x . Intuitivamente, la supervivencia en el período siguiente ($x+1$) será:

$$l_{x+1} = l_x - d_x \quad (7)$$

- Población a la edad x . Se representa con L_x , y son los años vividos por la población a la edad cumplida x . Equivale a:

$$L_x = l_{x+1} + d_x \text{ promedio de años vividos en la última edad } x \quad (8)$$

- Tanto específico de mortalidad a la edad x . Se representa con m_x . Es equivalente al número de personas que fallece con la edad cumplida x , en función del tiempo que llevaban expuestas al riesgo de mortalidad de ese período. En otras palabras, es el cociente entre los fallecimientos a la edad x y los años vividos por las personas de dicha edad, es decir:

$$m_x = \frac{d_x}{L_x} \quad (9)$$

- Probabilidad de fallecimiento en edad x . También llamada q_x , representa la probabilidad de que una persona de cierta edad x fallezca dentro de un año (o lo que es lo mismo, que muera con esa misma edad). Se utiliza para evaluar y predecir el riesgo de fallecimiento de una población específica, en un período de tiempo determinado. Se calcula mediante los datos históricos de fallecimiento de una población determinada, que se desglosan normalmente por edad, género y algunos factores relevantes. A partir de estos datos se puede calcular, por ejemplo, el número de fallecimientos de una población durante un período de tiempo sobre el tamaño medio de la población en ese período (lo que se denomina tasa bruta de mortalidad), o, como se mencionaba antes, la probabilidad de que una persona de edad x fallezca dentro de un año (es decir, la tasa de mortalidad para dicha edad x). En términos analíticos, sería el cociente entre el número de fallecimientos a la edad x y las personas expuestas al riesgo de mortalidad del período (es decir, los supervivientes a la edad x). Manteniendo el formato de formulación anterior, sería:

$$q_x = \frac{d_x}{l_x} \quad (10)$$

- Esperanza de vida. Se representa con e_x . Es el promedio de años que le quedarían por vivir a un individuo de edad x . Se obtiene como el cociente entre los años restantes de vida de los individuos una vez cumplida la edad x y el número de supervivientes con esa edad x :

$$e_x = \frac{\sum_{z \geq x} l_z}{l_x} \quad (11)$$

Por otro lado, para llevar a cabo los cálculos de estas métricas, se asume que la población está constituida por 100.000 personas y, a partir de ahí, se obtienen las exposiciones al riesgo de mortalidad y el resto de las funciones biométricas.

En el caso de España, las tablas de mortalidad se obtienen partiendo del número de defunciones que presenta el Movimiento Natural de la Población con fecha de 01 de enero de cada año. El movimiento Natural de la Población engloba los nacimientos, defunciones y matrimonios que suceden en España, partiendo de la información proporcionada por los Registros Civiles (INE, 2018).

4. RESULTADOS

Ahora que ya se han descrito los datos, se procede a la generación de los resultados. Se ha realizado un primer análisis para hombres y mujeres entre 0 y 90 años de edad. Sin embargo, por insuficiencia muestral, el análisis recogerá un rango de edades acotado entre los 11 y los 64 años. De este modo, se verificará que las tasas que afectan a los modelos son más significativas, a pesar de la falta de algunas edades.

4.1. Resultados GLM

Como se ha comentado, se han realizado diez modelos diferentes, cada uno representando un escenario distinto. Los primeros cinco modelos corresponden a los hombres, mientras que los últimos cinco son para mujeres. Cada uno de los modelos ha sido escogido mediante el criterio de información de Akaike (AIC), que es una medida de calidad relativa y se obtiene a través de la suma de los residuos cuadráticos de cada modelo. Entonces, se han realizado varios modelos y, finalmente, se ha optado por escoger el que tuviera un AIC menor, pues cuanto menor es el valor del criterio de información, mejor es el modelo que se utiliza.

El primer modelo representaría a los hombres que consumen tabaco diariamente, así como cannabis. Además, su Índice de Masa Corporal demuestra que tienen sobrepeso. También realizan ejercicio entre uno y dos días semanales, y duermen más de ocho horas diarias. La ecuación sería la siguiente:

$$Mortalidad_{hombres} = \beta_0 + \beta_1 Ejercicio_{1,2} + \beta_2 Sobrepeso + \beta_3 Sueño + \beta_4 Edad + \beta_5 Fumador_{diario} Ca + \varepsilon \quad (12)$$

Figura 4.1
Modelo 1

Coefficients:	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11041.255	897.666	12.300	< 0.0000000000000002	***
Hombres\$Ejercicio_1_2	-16745.011	3556.249	-4.709	0.00002158114872672	***
Hombres\$Sobrepeso	-19824.277	1640.182	-12.087	0.00000000000000036	***
Hombres\$Sueno	-2276.397	381.169	-5.972	0.00000027652890369	***
Hombres\$Edad	-17.532	1.719	-10.201	0.00000000000013129	***
Hombres\$Fumador_diario:Hombres\$Ca	25798.895	5576.747	4.626	0.00002842865969679	***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

Fuente: Elaboración propia

La figura 4.1 muestra los coeficientes estimados del primer modelo. Todos son significativos a niveles de confianza superiores al 99%. Se aprecia cómo el efecto de consumir cannabis y tabaco diario (que se relacionan como una interacción) favorece la mortalidad. Asimismo, el hecho de que tenga sobrepeso no implica algo negativo, pues es una persona que realiza actividad física durante uno o dos días semanales y descansa el tiempo necesario. Por esta misma razón, el efecto de la edad se reduce. El criterio de información de Akaike es de -763,27.

El segundo modelo representa a los hombres que ya no fuman, por tanto, exfumadores, pero que consumen hipnosedantes, no realizan ningún día a la semana ejercicio y su Índice de Masa Corporal indica que su peso es insuficiente. No obstante, descansan, al menos, ocho horas diarias. La ecuación sería:

$$Mortalidad_{hombres} = \beta_0 + \beta_1 Sueño + \beta_2 Edad + \beta_3 ExfumadorHip + \beta_4 Ejercicio_0 Insuficiente + \varepsilon \quad (13)$$

Figura 4.2
Modelo 2

Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2800.525	231.651	12.089	0.00000000000000257 ***
Hombres\$Sueno	-861.962	416.483	-2.070	0.0438 *
Hombres\$Edad	-20.955	4.914	-4.265	0.000091013400418976 ***
Hombres\$Exfumador:Hombres\$Hip	-22194.604	4222.841	-5.256	0.000003199095769921 ***
Hombres\$Ejercicio_0:Hombres\$Insuficiente	279960.203	54586.354	5.129	0.000004967012811953 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fuente: Elaboración propia

En este caso, el sueño pierde un poco de significatividad, aunque sigue siendo válido para el análisis si valoramos un nivel de confianza mayor al 95%. Sus efectos son beneficiosos para la salud. Una vez más, la edad no afecta de forma directa a la mortalidad. En cuanto a la variable exfumador, tiene efectos contrarios a la mortalidad, pero estos son menores a causa del consumo de hipnosedantes. El hecho de no hacer ejercicio y de que su peso sea insuficiente, afecta de forma negativa a la salud. El criterio de información de Akaike es de -659,23.

El tercer modelo, mostrado en la figura 4.3, se atribuye a hombres que han consumido alcohol en los últimos doce meses. Es un escenario mucho más genérico que los anteriores, en el que se ve cómo el alcoholismo afecta de forma favorable a la salud, como

ya se comentaba en la descripción de la variable. La ecuación del modelo tendría la siguiente forma:

$$Mortalidad_{hombres} = \beta_0 + \beta_1 Alcohol + \beta_2 Edad + \varepsilon \quad (14)$$

Figura 4.3
Modelo 3

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10450.49	1476.34	7.079	0.0000000409 ***
Hombres\$Alcohol	-10372.84	1876.54	-5.528	0.00000111558 ***
Hombres\$Edad	-36.26	2.60	-13.945	< 0.0000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fuente: Elaboración propia

Los efectos son significativos con un nivel de confianza superior al 99% y el criterio de información de Akaike tiene un valor de -673,64.

El cuarto modelo es representativo para hombres no fumadores, obesos, que se ejercitan entre tres y cuatro días semanales y con niveles de renta superiores al percentil 70 de la mediana de salarios en España. A continuación, se muestra la ecuación del modelo:

$$Mortalidad_{hombres} = \beta_0 + \beta_1 NoFumador + \beta_2 Renta + \beta_3 Edad + \beta_4 Ejercicio_{3,4} + \beta_5 Obesidad + \varepsilon \quad (15)$$

Figura 4.4
Modelo 4

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12388.535	1852.157	6.689	0.0000000201 ***
Hombres\$No_fumador	-10111.010	1940.416	-5.211	0.0000037407 ***
Hombres\$Renta	2355.507	379.815	6.202	0.0000001143 ***
Hombres\$Edad	-17.956	5.133	-3.498	0.00101 **
Hombres\$Ejercicio_3_4:Hombres\$Obesidad	-202027.779	32609.158	-6.195	0.0000001169 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fuente: Elaboración propia

En este caso, el efecto de la renta alta tiene una influencia negativa para la salud, siendo favorable a la mortalidad. En cuanto a la variable de no tabaquismo, es beneficiosa, así como la interacción entre la variable obesidad y las tasas de ejercicio físico semanal. A pesar de presentar un Índice de Masa Corporal muy elevado, su influencia se ve mermada

por la actividad física recurrente. El valor del criterio de información de Akaike es de -684,10.

El quinto y último modelo para los hombres muestra la influencia del tabaquismo ocasional, el consumo de hipnosedantes, el descanso superior a ocho horas diarias y la ejercitación física entre cinco y seis días semanales de un varón cuyo Índice de Masa Corporal es normal, es decir, está incluido en el rango de normopeso. Su ecuación sería:

$$Mortalidad_{hombres} = \beta_0 + \beta_1 Fumador_{ocasional} + \beta_2 Hip + \beta_3 Ejercicio_{5,6} + \beta_4 Normopeso + \beta_5 Sueño + \beta_6 Edad + \varepsilon \quad (16)$$

Figura 4.5
Modelo 5

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	11381.94	1039.42	10.950	0.00000000000001578	***
Hombres\$Fumador_ocasional	27695.34	7081.58	3.911	0.000295	***
Hombres\$Hip	-48193.05	4019.74	-11.989	0.00000000000000067	***
Hombres\$Ejercicio_5_6	33632.39	4143.78	8.116	0.00000000017085350	***
Hombres\$Normopeso	-4296.18	1169.23	-3.674	0.000610	***
Hombres\$Sueno	-5684.08	590.97	-9.618	0.00000000000110791	***
Hombres\$Edad	-17.09	2.48	-6.892	0.00000001191487020	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fuente: Elaboración propia

Como así lo muestra la figura 4.5, todos los coeficientes son significativos con una confianza mayor al 99%. Las tasas de tabaquismo ocasional están muy relacionadas con la mortalidad. Para una persona con una complejión favorable, la realización de más de cinco días de ejercicio físico semanal tiene efectos adversos. El consumo de hipnosedantes tiene efectos negativos para la mortalidad. Finalmente, el coeficiente para la edad es negativo, pues favorece a la salud por el hecho de ir acompañado de unas horas de sueño superiores a las recomendadas. Para finalizar, el criterio de información de Akaike es de -761,29.

Ahora se describirán los escenarios para la población femenina de España. Los modelos son diferentes a los de los hombres, puesto que los factores de riesgo no afectan de la misma manera a ambas secciones de la población. Al igual que con los hombres, se han propuesto cinco modelos para explicar posibles escenarios futuros. Se numerarán a partir del número seis, para seguir con el orden establecido en la modelización.

El sexto modelo representa a las mujeres que consumían tabaco en algún momento pasado, pero que, en la fecha de realización de la encuesta, no lo hacían. Además, presentan un nivel de renta superior al percentil 70 de los salarios medianos en España (como recordatorio, se ha asumido que son iguales que los de los hombres) y tienen el Índice de Masa Corporal más elevado, aunque realizan ejercicio físico como máximo dos días a la semana. Su ecuación tendría la siguiente forma:

$$Mortalidad_{mujeres} = \beta_0 + \beta_1 Exfumador + \beta_2 Renta + \beta_3 Edad + \beta_4 Ejercicio_{1,2} Obesidad + \varepsilon \quad (17)$$

Figura 4.6
Modelo 6

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14330.68	1683.56	8.512	0.0000000000318 ***
Mujeres\$Exfumador	77057.04	14295.50	5.390	0.0000020034852 ***
Mujeres\$Renta	-5618.13	1071.90	-5.241	0.0000033650902 ***
Mujeres\$Edad	-35.77	10.61	-3.370	0.00147 **
Mujeres\$Ejercicio_1_2:Mujeres\$Obesidad	-1295228.78	218396.14	-5.931	0.0000002992770 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Fuente: Elaboración propia

Los resultados demuestran que, a pesar de no consumir tabaco actualmente, sus efectos son adversos para la salud. Por el contrario, los niveles superiores de renta son positivos, así como la interacción entre el Índice de Masa Corporal y el ejercicio físico. Dadas las características, la edad mantiene efectos reductores en la mortalidad. El criterio de información de Akaike tiene un valor de -759,96.

El séptimo modelo representa a las mujeres que consumen tabaco de forma ocasional y también han ingerido alcohol en los últimos doce meses. Su Índice de Masa Corporal indica que tienen un peso saludable, sin embargo, no realizan actividad física ningún día por semana. La ecuación sería:

$$Mortalidad_{mujeres} = \beta_0 + \beta_1 Edad + \beta_2 Fumador_{ocasional} Alcohol + \beta_3 Ejercicio_0 Normopeso + \varepsilon \quad (18)$$

Figura 4.7
Modelo 7

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6021.457	1076.665	5.593	0.0000009333877767	***
Mujeres\$Edad	-88.345	8.514	-10.376	0.00000000000000457	***
Mujeres\$Fumador_ocasional:Mujeres\$Alcohol	228917.958	50483.623	4.534	0.0000362784620029	***
Mujeres\$Ejercicio_0:Mujeres\$Normopeso	-13615.110	2357.213	-5.776	0.0000004869825451	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fuente: Elaboración propia

Los resultados sugieren que, a pesar de no realizar nada de ejercicio, el hecho de mantener una relación saludable con el peso es beneficioso para la salud. Por otro lado, el consumo ocasional de tabaco acompañado de alcohol provoca efectos que perjudican a la salud. El criterio de información de Akaike es de -741,43.

La figura 4.8 muestra el modelo número ocho. Este escenifica a una mujer que consume tabaco ocasionalmente, así como alcohol y cannabis. También realiza ejercicio físico hasta dos días semanales y tiene sobrepeso. La ecuación sería la siguiente:

$$Mortalidad_{mujeres} = \beta_0 + \beta_1 Fumador_{ocasional} + \beta_2 Alcohol + \beta_3 Can + \beta_4 Edad + \beta_5 Ejercicio_{1,2} Sobrepeso + \varepsilon \quad (19)$$

Figura 4.8
Modelo 8

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	14487.72	1416.17	10.230	0.000000000000012	***
Mujeres\$Fumador_ocasional	58593.13	23329.48	2.512	0.0154	*
Mujeres\$Alcohol	-24323.56	1975.72	-12.311	< 0.0000000000000002	***
Mujeres\$Can	16840.24	3122.51	5.393	0.00000208783539	***
Mujeres\$Edad	-34.66	5.50	-6.302	0.00000008646122	***
Mujeres\$Ejercicio_1_2:Mujeres\$Sobrepeso	-85395.26	9924.54	-8.604	0.00000000002730	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fuente: Elaboración propia

Todos los coeficientes son significativos a un nivel de confianza, como mínimo, del 95%. Las tasas de tabaquismo ocasional favorecen a la mortalidad, así como el consumo de cannabis. En cuanto al alcohol, tiene un efecto negativo en términos de mortalidad, ya que se relaciona de manera independiente con el tabaco y el cannabis, es decir, que los efectos del consumo de alcohol no dependen de los efectos de las otras dos variables. En cuanto al sobrepeso, este no afecta adversamente a la salud porque realiza actividad física semanal. El criterio de información de Akaike es de -830,24.

A continuación, se encuentra el modelo nueve, que representa a las mujeres que ya no fuman, es decir, que son exfumadoras, y que realizan ejercicio físico hasta seis días por semana, con un Índice de Masa Corporal inferior al recomendable. Su ecuación se muestra a continuación:

$$Mortalidad_{mujeres} = \beta_0 + \beta_1 Exfumador + \beta_2 Edad + \beta_3 Ejercicio_{5,6} Insuficiente + \varepsilon \quad (20)$$

Figura 4.9
Modelo 9

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4065.50	359.47	11.310	0.0000000000000218	***
Mujeres\$Exfumador	-9584.84	2093.10	-4.579	0.00003121624994011	***
Mujeres\$Edad	-36.35	9.14	-3.978	0.000225	***
Mujeres\$Ejercicio_5_6:Mujeres\$Insuficiente	468015.03	81038.30	5.775	0.00000048820321557	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fuente: Elaboración propia

A diferencia de lo sucedido en el modelo seis, las tasas de tabaquismo son favorables a la salud. No obstante, al mantener un peso insuficiente, la realización de actividad física tan intensa afecta de forma negativa a la salud. El criterio de información de Akaike es de -741,18.

Finalmente, se muestra el último modelo. En él, únicamente se evalúa el efecto de la realización de ejercicio físico entre tres y cuatro días semanales, dependiendo de la edad de las mujeres. La ecuación del modelo tendría la siguiente forma:

$$Mortalidad_{mujeres} = \beta_0 + \beta_1 Ejercicio_{3,4} Edad + \varepsilon \quad (21)$$

Figura 4.10
Modelo 10

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5898.60	593.00	9.947	0.0000000000000125	***
Mujeres\$Ejercicio_3_4:Mujeres\$Edad	-545.99	56.36	-9.688	0.0000000000000307	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fuente: Elaboración propia

La figura 4.10 muestra los resultados del décimo modelo. El supuesto de que únicamente se realice ejercicio semanal entre tres y cuatro días, y que esto dependa de la edad, implica unos efectos positivos para la salud. El criterio de información de Akaike tiene un valor de -702,54.

A continuación, se muestra la tabla 4.1, con las predicciones de las tasas de mortalidad para cada modelo, influidas por los factores de riesgo correspondientes.

Tabla 4.1
Predicción de las tasas de mortalidad para cada modelo

Edades	Predicción modelo 1	Predicción modelo 2	Predicción modelo 3	Predicción modelo 4	Predicción modelo 5	Predicción modelo 6	Predicción modelo 7	Predicción modelo 8	Predicción modelo 9	Predicción modelo 10
11	0,00092	0,00389	0,00099	0,00082	0,00089	0,00072	0,00198	0,00071	0,00273	0,00170
12	0,00092	0,00392	0,00100	0,00082	0,00089	0,00072	0,00202	0,00071	0,00276	0,00170
13	0,00092	0,00396	0,00100	0,00082	0,00090	0,00072	0,00205	0,00071	0,00278	0,00170
14	0,00093	0,00399	0,00101	0,00082	0,00090	0,00072	0,00209	0,00071	0,00281	0,00170
15	0,00100	0,00425	0,00293	0,00215	0,00084	0,00061	0,00101	0,00103	0,00314	0,00251
16	0,00130	0,00681	0,00296	0,00216	0,00162	0,00061	0,00102	0,00103	0,00318	0,00259
17	0,00131	0,00691	0,00300	0,00217	0,00162	0,00061	0,00103	0,00104	0,00322	0,00268
18	0,00332	0,00282	0,00303	0,00687	0,00313	0,00244	0,00174	0,00150	0,00140	0,00277
19	0,00334	0,00283	0,00306	0,00696	0,00315	0,00246	0,00177	0,00151	0,00140	0,00288
20	0,00336	0,00285	0,00310	0,00705	0,00317	0,00248	0,00180	0,00152	0,00141	0,00299
21	0,00338	0,00287	0,00313	0,00714	0,00319	0,00250	0,00183	0,00153	0,00142	0,00310
22	0,00340	0,00289	0,00317	0,00723	0,00320	0,00253	0,00186	0,00154	0,00143	0,00323
23	0,00342	0,00290	0,00321	0,00732	0,00322	0,00255	0,00189	0,00154	0,00143	0,00337
24	0,00344	0,00292	0,00324	0,00742	0,00324	0,00257	0,00192	0,00155	0,00144	0,00352
25	0,00494	0,00421	0,00845	0,00523	0,00501	0,00278	0,00271	0,00207	0,00196	0,00324
26	0,00498	0,00425	0,00872	0,00528	0,00505	0,00281	0,00278	0,00208	0,00197	0,00337
27	0,00503	0,00429	0,00900	0,00533	0,00510	0,00284	0,00285	0,00210	0,00199	0,00350
28	0,00507	0,00433	0,00931	0,00538	0,00514	0,00287	0,00292	0,00211	0,00200	0,00364
29	0,00512	0,00437	0,00963	0,00543	0,00519	0,00290	0,00300	0,00213	0,00202	0,00380
30	0,00516	0,00441	0,00998	0,00549	0,00524	0,00293	0,00308	0,00214	0,00203	0,00397
31	0,00521	0,00445	0,01035	0,00554	0,00528	0,00296	0,00316	0,00216	0,00205	0,00415
32	0,00526	0,00449	0,01076	0,00560	0,00533	0,00299	0,00326	0,00217	0,00206	0,00436
33	0,00531	0,00453	0,01119	0,00565	0,00538	0,00302	0,00335	0,00219	0,00208	0,00458
34	0,00536	0,00458	0,01167	0,00571	0,00543	0,00306	0,00345	0,00221	0,00209	0,00483
35	0,00863	0,00797	0,00875	0,00853	0,00885	0,00486	0,00379	0,00468	0,00470	0,00511
36	0,00876	0,00811	0,00904	0,00866	0,00898	0,00495	0,00392	0,00476	0,00478	0,00542
37	0,00890	0,00825	0,00935	0,00880	0,00912	0,00504	0,00406	0,00484	0,00487	0,00577
38	0,00904	0,00839	0,00968	0,00894	0,00927	0,00513	0,00422	0,00492	0,00496	0,00617
39	0,00918	0,00854	0,01003	0,00908	0,00942	0,00523	0,00438	0,00500	0,00505	0,00663
40	0,00933	0,00870	0,01041	0,00924	0,00957	0,00532	0,00456	0,00509	0,00514	0,00717
41	0,00949	0,00886	0,01081	0,00939	0,00973	0,00543	0,00475	0,00518	0,00524	0,00780
42	0,00965	0,00903	0,01125	0,00955	0,00989	0,00554	0,00495	0,00528	0,00534	0,00855
43	0,00982	0,00920	0,01173	0,00972	0,01006	0,00565	0,00518	0,00538	0,00545	0,00946
44	0,00999	0,00938	0,01226	0,00989	0,01024	0,00576	0,00543	0,00548	0,00556	0,01058
45	0,00297	0,01931	0,01455	0,01558	0,002275	0,00933	0,00892	0,01229	0,00862	0,00558
46	0,00293	0,02012	0,01536	0,01603	0,002367	0,00965	0,00968	0,01284	0,00890	0,00588
47	0,002498	0,02101	0,01626	0,01650	0,002467	0,01000	0,01059	0,01344	0,00920	0,00621
48	0,002612	0,02198	0,01728	0,01701	0,002575	0,01037	0,01168	0,01409	0,00952	0,00658
49	0,002738	0,02304	0,01844	0,01754	0,002694	0,01077	0,01302	0,01482	0,00986	0,00700
50	0,002876	0,02421	0,01976	0,01811	0,002824	0,01120	0,01471	0,01562	0,01023	0,00748
51	0,003028	0,02550	0,02128	0,01872	0,002967	0,01167	0,01691	0,01651	0,01062	0,00803
52	0,003198	0,02694	0,02306	0,01937	0,003126	0,01218	0,01988	0,01751	0,01105	0,00866
53	0,003388	0,02855	0,02517	0,02007	0,003302	0,01273	0,02412	0,01865	0,01151	0,00941
54	0,003602	0,03037	0,02770	0,02082	0,003500	0,01334	0,03065	0,01993	0,01201	0,01029
55	0,004550	0,04176	0,02785	0,04666	0,004620	0,02244	0,01152	0,02203	0,02253	0,01136
56	0,004944	0,04577	0,03098	0,05092	0,005016	0,02439	0,01283	0,02385	0,02454	0,01267
57	0,005413	0,05062	0,03491	0,05605	0,005486	0,02673	0,01447	0,02600	0,02694	0,01432
58	0,005981	0,05663	0,03997	0,06232	0,006054	0,02955	0,01658	0,02857	0,02987	0,01648
59	0,006681	0,06426	0,04674	0,07017	0,006753	0,03304	0,01943	0,03172	0,03351	0,01939
60	0,007568	0,07425	0,05628	0,08028	0,007634	0,03747	0,02346	0,03563	0,03815	0,02356
61	0,008726	0,08794	0,07071	0,09381	0,008780	0,04328	0,02959	0,04065	0,04430	0,03001
62	0,010302	0,10780	0,09509	0,11281	0,010330	0,05120	0,04007	0,04732	0,05280	0,04133
63	0,012572	0,13926	0,14513	0,14147	0,012545	0,06268	0,06202	0,05661	0,06535	0,06634
64	0,016127	0,19665	0,03638	0,18964	0,015969	0,08080	0,013718	0,07043	0,08571	0,16803

Fuente: Elaboración propia

Como resumen de los resultados de los modelos GLM, se podría decir que, para los primeros años de la serie, es decir, entre los 11 y los 37 años, el modelo que proporciona una menor mortalidad es el número uno para los hombres y, para las mujeres, el número ocho. Por otro lado, para el segundo tramo de la serie, es decir, entre 38 y 64 años, la menor mortalidad se atribuye al modelo número tres de los hombres y el diez para las mujeres. Esto quiere decir que, entre los hipotéticos escenarios, sería más saludable el hombre entre 11 y 37 años que fume a diario, consuma cannabis, se ejercite uno o dos días por semana, tenga sobrepeso y duerma las horas de sueño necesarias. En ese mismo contexto, la mujer más saludable sería la que fume ocasionalmente, haya consumido alcohol y cannabis en el último año, se ejercite entre una y dos veces por semana y tenga

sobrepeso. Para el segundo tramo de la serie, el hombre más saludable sería aquel que únicamente haya consumido alcohol en los últimos doce meses y la mujer más saludable, la que se ejercite entre tres y cuatro días a la semana.

4.2. Resultados Lee-Carter

El objetivo del análisis es obtener las proyecciones para el año 2030 de las tasas de mortalidad influidas por los diferentes factores de riesgo y compararlas con las tasas de mortalidad reales proyectadas para ese mismo año. Para ello, se aplicará el modelo de Lee-Carter explicado previamente, tanto para los resultados de los diferentes modelos, como para las tasas de mortalidad reales de hombres y mujeres.

Como se venía comentando, el período del análisis está comprendido entre los años 1960 y 2020, y las edades de estudio, entre los 11 y los 64 años.

En primer lugar, se mostrarán los resultados de las tasas de mortalidad reales proyectadas para las mujeres.

Figura 4.11

Resultados de las tasas de mortalidad reales proyectadas para las mujeres

```
Adjustment method: dt
Region: Spain
Years in fit: 1960 - 2020
Ages in fit: 11 - 64

Percentage variation explained: 92.3%

ERROR MEASURES BASED ON MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00000 0.00000 0.01695 0.10423

Averages across years:
      IE      ISE      IPE      IAPE
0.00010 0.00000 0.89522 5.40474

ERROR MEASURES BASED ON LOG MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00645 0.02093 -0.00064 0.01403

Averages across years:
      IE      ISE      IPE      IAPE
0.34104 1.05200 -0.03356 0.72747
```

Fuente: Elaboración propia

La figura 4.11 muestra el ajuste de Lee-Carter para las tasas de mortalidad de las mujeres. El modelo utilizado para hacer la proyección es correcto, ya que explica el 92,3% de la variabilidad, lo cual es muy positivo.

Figura 4.12

Ajuste Lee-Carter para las tasas de mortalidad reales proyectadas para las mujeres

```
Years: 2021 - 2030
Ages: 11 - 64

ERROR MEASURES BASED ON MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00000 0.00000 0.01695 0.10423

Averages across years:
      IE      ISE      IPE      IAPE
0.00010 0.00000 0.89522 5.40474

ERROR MEASURES BASED ON LOG MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00645 0.02093 -0.00064 0.01403

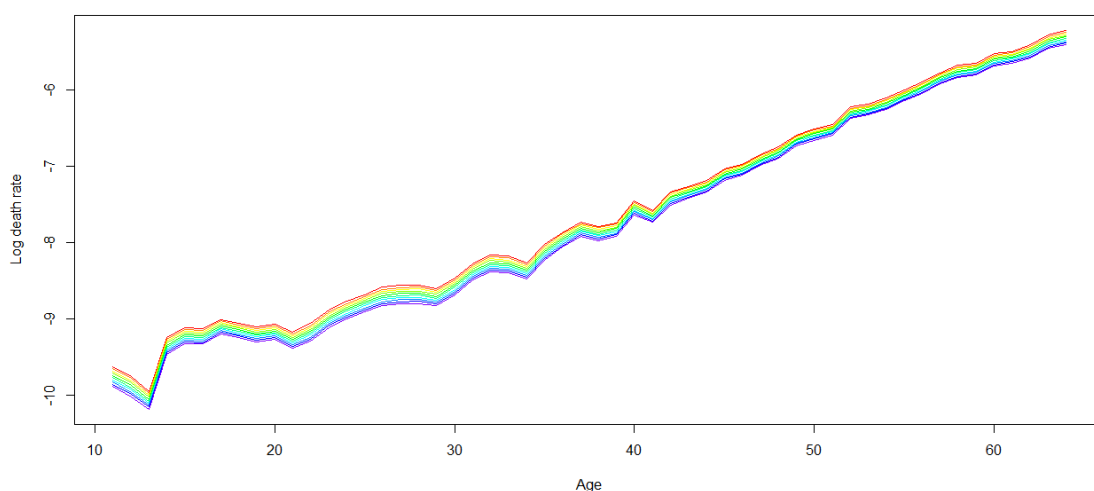
Averages across years:
      IE      ISE      IPE      IAPE
0.34104 1.05200 -0.03356 0.72747
```

Fuente: Elaboración propia

La figura 4.12 muestra las proyecciones de estas tasas de mortalidad en diez años. Se han incluido, además, métricas como el error medio¹, el error cuadrático medio², el error integrado³ o el error porcentual medio⁴. Si se centra la atención en las medidas ME, MSE, IE e ISE⁵, son muy favorables, ya que los errores son nulos. Esto implica que las predicciones no tienen sesgo sistemático significativo. En cuanto a el MPE y el IPE⁶, los resultados sugieren que las proyecciones están muy cerca de los valores reales.

Figura 4.13

Proyección de las tasas de mortalidad reales para las mujeres, 2021 - 2030



Fuente: Elaboración propia

La figura 4.13 grafica las tasas proyectadas obtenidas. El modelo Lee-Carter presenta las proyecciones en formato logarítmico. Cada una de las líneas tiene un color diferente, que representa cada uno de los diez años proyectados de las tasas de mortalidad reales de las mujeres, es decir, la línea morada muestra la evolución de las tasas de mortalidad para el año 2021, mientras que la roja lo hace para el año 2030. En el siguiente epígrafe se analizará más profundamente este resultado.

¹ ME

² MSE

³ IE

⁴ MPE

⁵ Error cuadrático integrado

⁶ Error porcentual integrado

Ahora se presentarán los resultados de las tasas de mortalidad reales proyectadas para los hombres.

Figura 4.14

Resultados de las tasas de mortalidad reales proyectadas para los hombres

```

Adjustment method: dt
Region: Spain
Years in fit: 1960 - 2020
Ages in fit: 11 - 64

Percentage variation explained: 90.2%

ERROR MEASURES BASED ON MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00000 0.00000 0.01620 0.10425

Averages across years:
      IE      ISE      IPE      IAPE
-0.00003 0.00001 0.85587 5.43357

ERROR MEASURES BASED ON LOG MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00587 0.02113 -0.00052 0.01632

Averages across years:
      IE      ISE      IPE      IAPE
0.31058 1.08541 -0.02778 0.85253

```

Fuente: Elaboración propia

La figura 4.14 muestra el ajuste de Lee-Carter para las tasas de mortalidad de los hombres. En este caso, el modelo explica el 90,2% de la variabilidad. Aunque es ligeramente más débil que el modelo de las mujeres, sigue siendo muy correcto para predecir la mortalidad.

Figura 4.15

Ajuste Lee-Carter para las tasas de mortalidad reales proyectadas para los hombres

```
Years: 2021 - 2030
Ages: 11 - 64

ERROR MEASURES BASED ON MORTALITY RATES

Averages across ages:
  ME      MSE      MPE      MAPE
0.00000  0.00000  0.01620  0.10425

Averages across years:
  IE      ISE      IPE      IAPE
-0.00003  0.00001  0.85587  5.43357

ERROR MEASURES BASED ON LOG MORTALITY RATES

Averages across ages:
  ME      MSE      MPE      MAPE
0.00587  0.02113  -0.00052  0.01632

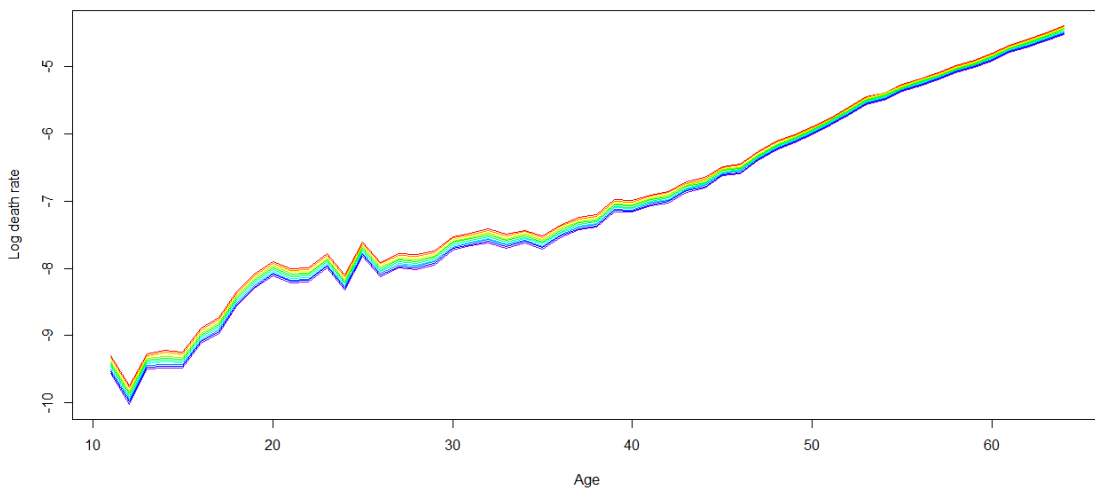
Averages across years:
  IE      ISE      IPE      IAPE
0.31058  1.08541  -0.02778  0.85253
```

Fuente: Elaboración propia

La figura 4.15 muestra los resultados de las proyecciones dentro de diez años de las tasas de mortalidad de los hombres. Al igual que ocurría con los resultados de las mujeres, las métricas son muy favorables.

Figura 4.16

Proyección de las tasas de mortalidad reales para los hombres, 2021 - 2030



Fuente: Elaboración propia

La figura 4.16 es un gráfico de las tasas anteriores en formato logarítmico. Al igual que ocurría con el gráfico de las mujeres, este muestra la evolución de las tasas de mortalidad reales proyectadas para los hombres, siendo la línea morada el año 2021 y la roja, el año 2030.

Además de observar el porcentaje de variabilidad que explica cada modelo de Lee-Carter, se ha decidido realizar un backtesting, con el objetivo de verificar que los modelos que se utilicen sean lo más precisos posible. Para llevarlo a cabo, se ha dividido la muestra en train set, incluyendo los años 1960 hasta 2010, y en test set, que incluye los años entre 2011 hasta 2020. Con este train set se ha tratado de hacer una predicción de las tasas de mortalidad hasta el año 2020, con la finalidad de compararlas con las tasas reales de ese año y comprobar la bondad del ajuste. Los resultados son los expuestos en las figuras 4.17 y 4.18.

Figura 4.17
Resultados del train set

```

Adjustment method: dt
Region: Spain
Years in fit: 1960 - 2010
Ages in fit: 11 - 64

Percentage variation explained: 85.2%

ERROR MEASURES BASED ON MORTALITY RATES

Averages across ages:
      ME      MSE      MPE      MAPE
0.00000 0.00000 0.01190 0.08091

Averages across years:
      IE      ISE      IPE      IAPE
0.00016 0.00000 0.62993 4.24352

ERROR MEASURES BASED ON LOG MORTALITY RATES

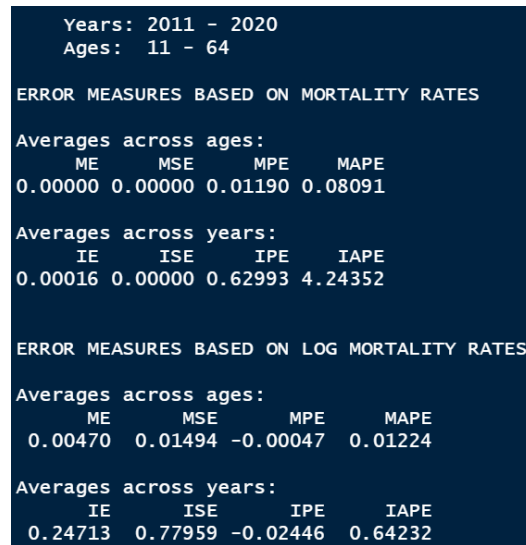
Averages across ages:
      ME      MSE      MPE      MAPE
0.00470 0.01494 -0.00047 0.01224

Averages across years:
      IE      ISE      IPE      IAPE
0.24713 0.77959 -0.02446 0.64232

```

Fuente: Elaboración propia

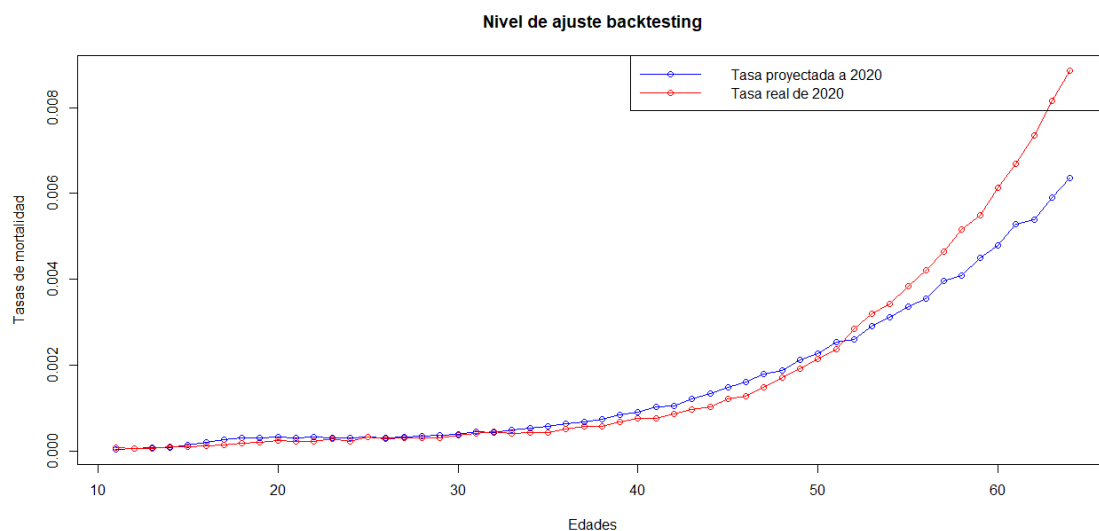
Figura 4.18
Resultados del test set



Fuente: Elaboración propia

Estos resultados del backtesting reflejan que el modelo utilizado para hacer las predicciones es correcto, pues explica el 85,2% de la variabilidad de los datos. Además, las métricas de error son muy bajas, por lo que se refuerza el concepto de la idoneidad del modelo.

Figura 4.19
Ajuste del backtesting con respecto a las tasas de mortalidad reales



Fuente: Elaboración propia

La figura 4.19 muestra el nivel de ajuste del backtesting explicado, a través de la representación de las tasas reales de mortalidad del año 2020 (en color rojo) y la proyección de las tasas del train set en el año 2020 (en color azul). El ajuste es muy preciso, sobre todo durante los primeros años de la serie y a medida que avanza la edad, las tasas comienzan a presentar algunas diferencias. Esto indicaría que el modelo es más fiable para predecir las edades menores, pero algo peor para las edades finales de la serie. No obstante, al tratarse de valores tan cercanos a 0, las diferencias son ínfimas.

5. PROYECCIONES DE MORTALIDAD: 2020 VS. 2030

Ahora se van a mostrar las proyecciones de los modelos y de las tasas reales de mortalidad para los hombres y mujeres en el año 2030.

Tabla 5.1

Proyección de las tasas de mortalidad reales y por modelo para los hombres, 2030

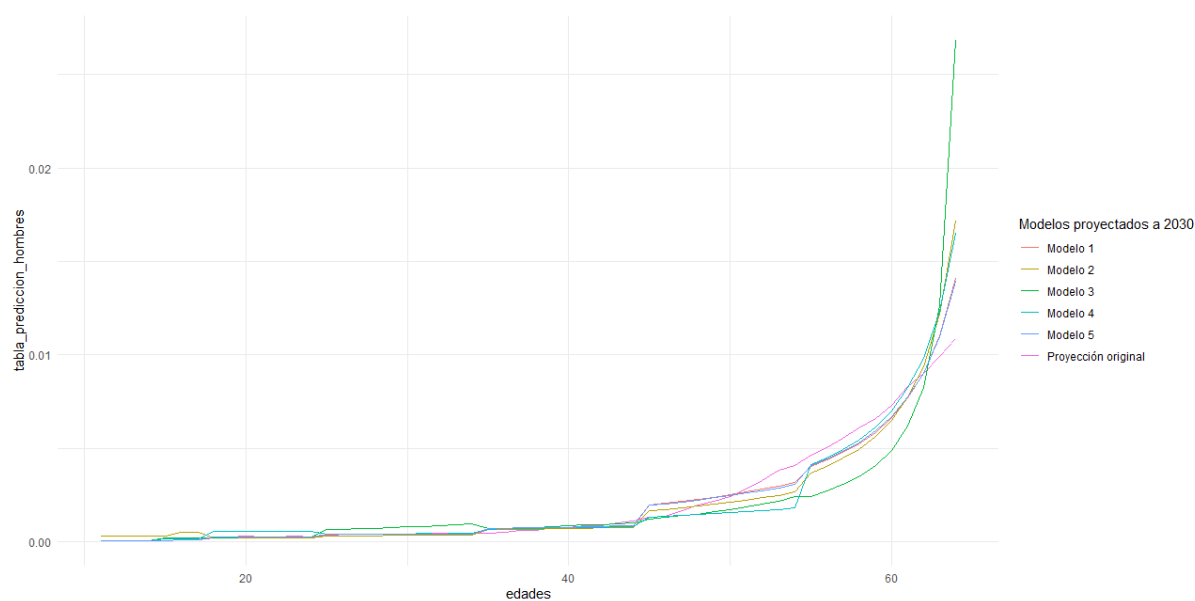
Edades	Tabla prediccion hombres	Tabla prediccion modelo 1	Tabla prediccion modelo 2	Tabla prediccion modelo 3	Tabla prediccion modelo 4	Tabla prediccion modelo 5
11	0,000070	0,000069	0,000293	0,000073	0,000060	0,000066
12	0,000045	0,000069	0,000296	0,000074	0,000061	0,000067
13	0,000075	0,000071	0,000309	0,000076	0,000063	0,000069
14	0,000076	0,000069	0,000300	0,000074	0,000061	0,000067
15	0,000076	0,000077	0,000332	0,000227	0,000166	0,000065
16	0,000110	0,000101	0,000537	0,000230	0,000167	0,000126
17	0,000127	0,000100	0,000537	0,000229	0,000166	0,000124
18	0,000190	0,000261	0,000220	0,000235	0,000542	0,000246
19	0,000248	0,000261	0,000220	0,000237	0,000545	0,000246
20	0,000299	0,000264	0,000222	0,000241	0,000556	0,000249
21	0,000269	0,000266	0,000224	0,000244	0,000563	0,000250
22	0,000273	0,000269	0,000227	0,000249	0,000576	0,000254
23	0,000337	0,000270	0,000228	0,000250	0,000580	0,000254
24	0,000243	0,000269	0,000227	0,000251	0,000583	0,000253
25	0,000400	0,000388	0,000329	0,000663	0,000409	0,000394
26	0,000297	0,000394	0,000334	0,000689	0,000415	0,000400
27	0,000336	0,000396	0,000336	0,000709	0,000418	0,000402
28	0,000329	0,000398	0,000337	0,000730	0,000420	0,000404
29	0,000352	0,000407	0,000345	0,000766	0,000430	0,000413
30	0,000439	0,000410	0,000348	0,000793	0,000434	0,000416
31	0,000467	0,000419	0,000356	0,000834	0,000444	0,000425
32	0,000490	0,000415	0,000352	0,000851	0,000440	0,000421
33	0,000453	0,000423	0,000359	0,000894	0,000449	0,000429
34	0,000488	0,000432	0,000367	0,000944	0,000459	0,000438
35	0,000446	0,000701	0,000643	0,000705	0,000688	0,000719
36	0,000529	0,000713	0,000656	0,000730	0,000701	0,000731
37	0,000594	0,000725	0,000668	0,000756	0,000712	0,000743
38	0,000617	0,000737	0,000681	0,000783	0,000724	0,000756
39	0,000780	0,000757	0,000700	0,000821	0,000744	0,000776
40	0,000771	0,000769	0,000714	0,000853	0,000757	0,000789
41	0,000850	0,000794	0,000739	0,000901	0,000782	0,000814
42	0,000887	0,000798	0,000743	0,000926	0,000786	0,000818
43	0,001041	0,000828	0,000774	0,000986	0,000816	0,000849
44	0,001108	0,000838	0,000785	0,001025	0,000827	0,000860
45	0,001326	0,001972	0,001648	0,001233	0,001324	0,001953
46	0,001373	0,002061	0,001724	0,001307	0,001367	0,002038
47	0,001662	0,002170	0,001816	0,001397	0,001419	0,002142
48	0,001955	0,002273	0,001903	0,001487	0,001465	0,002240
49	0,002157	0,002397	0,002008	0,001598	0,001521	0,002358
50	0,002449	0,002514	0,002108	0,001711	0,001568	0,002469
51	0,002830	0,002682	0,002250	0,001868	0,001642	0,002627
52	0,003270	0,002813	0,002361	0,002011	0,001687	0,002748
53	0,003845	0,002972	0,002496	0,002190	0,001743	0,002896
54	0,004104	0,003179	0,002671	0,002426	0,001819	0,003088
55	0,004650	0,004027	0,003686	0,002438	0,004118	0,004090
56	0,005053	0,004364	0,004029	0,002706	0,004483	0,004429
57	0,005522	0,004791	0,004469	0,003057	0,004947	0,004856
58	0,006120	0,005253	0,004961	0,003474	0,005458	0,005318
59	0,006624	0,005869	0,005630	0,004066	0,006147	0,005932
60	0,007323	0,006634	0,006495	0,004889	0,007020	0,006693
61	0,008305	0,007703	0,007747	0,006194	0,008262	0,007751
62	0,009035	0,008989	0,009388	0,008242	0,009818	0,009014
63	0,009940	0,011003	0,012168	0,012652	0,012351	0,010979
64	0,010886	0,014073	0,017143	0,026795	0,016512	0,013933

Fuente: Elaboración propia

En primer lugar, la tabla 5.1 muestra la información para los hombres. La primera columna representa la proyección de las tasas reales en el año 2030, mientras que las siguientes son las tasas proyectadas de cada uno de los cinco GLM.

Figura 5.1

Proyección de las tasas de mortalidad reales y por modelo para los hombres, 2030



Fuente: Elaboración propia

La figura 5.1 muestra el gráfico de las tasas reflejadas en la tabla 5.1. Aquí se ve más claramente el comportamiento de los distintos modelos con las tasas reales, representadas con el nombre “proyección original”. Todos los modelos fluctúan a lo largo del tiempo. En concreto, las tasas reales representan una mortalidad inferior al resto de modelos durante las primeras edades de la serie, es decir, entre los 11 y los 40 años aproximadamente. Más adelante, entre los 50 y los 60, las tasas reales son mayores al resto y, a partir de esa edad, los modelos presentan una mortalidad mucho mayor. Esto implica que, si se realizase un análisis para más edades con estas mismas características, los modelos presentarían tasas de mortalidad aún mayores que las reales.

Tabla 5.2

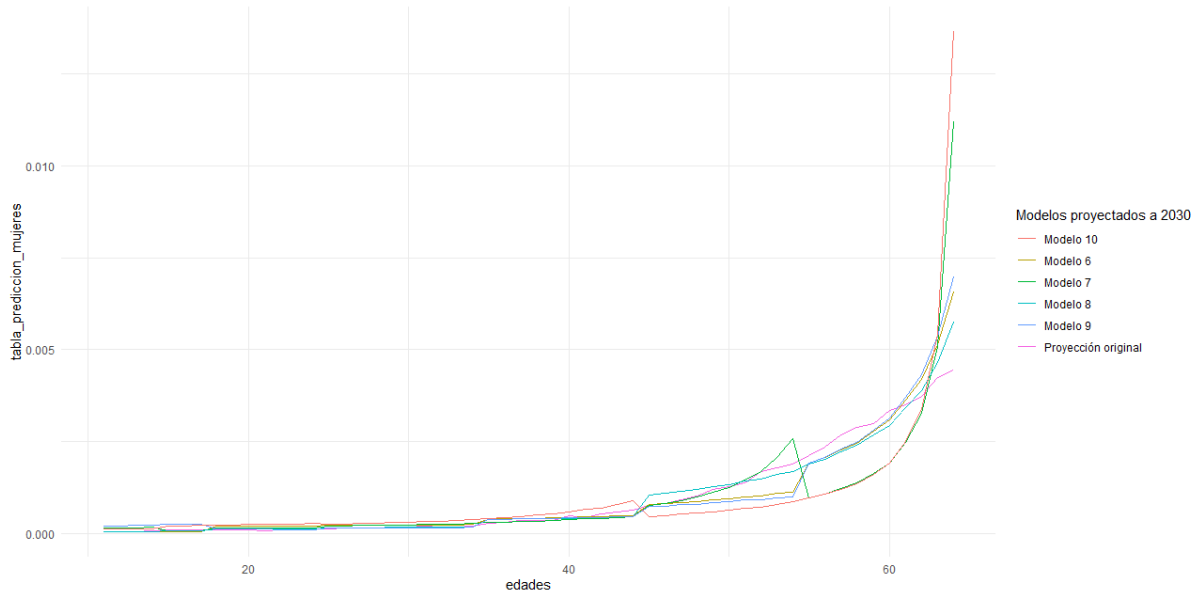
Proyección de las tasas de mortalidad reales y por modelo para las mujeres, 2030

Edades	Tabla prediccion mujeres	Tabla prediccion modelo 6	Tabla prediccion modelo 7	Tabla prediccion modelo 8	Tabla prediccion modelo 9	Tabla prediccion modelo 10
11	0,000051	0,000054	0,000149	0,000053	0,000206	0,000126
12	0,000045	0,000053	0,000151	0,000053	0,000207	0,000125
13	0,000038	0,000056	0,000161	0,000055	0,000219	0,000131
14	0,000078	0,000056	0,000164	0,000056	0,000221	0,000132
15	0,000089	0,000047	0,000080	0,000081	0,000250	0,000197
16	0,000089	0,000048	0,000081	0,000082	0,000255	0,000206
17	0,000102	0,000049	0,000084	0,000084	0,000263	0,000218
18	0,000097	0,000199	0,000141	0,000122	0,000113	0,000225
19	0,000092	0,000201	0,000144	0,000123	0,000114	0,000233
20	0,000095	0,000202	0,000146	0,000123	0,000114	0,000241
21	0,000084	0,000200	0,000146	0,000122	0,000113	0,000247
22	0,000093	0,000197	0,000144	0,000120	0,000110	0,000251
23	0,000111	0,000199	0,000147	0,000120	0,000111	0,000261
24	0,000123	0,000199	0,000148	0,000120	0,000111	0,000271
25	0,000136	0,000217	0,000211	0,000161	0,000152	0,000251
26	0,000146	0,000214	0,000211	0,000158	0,000149	0,000254
27	0,000151	0,000218	0,000218	0,000161	0,000152	0,000267
28	0,000151	0,000219	0,000223	0,000161	0,000152	0,000276
29	0,000146	0,000226	0,000233	0,000166	0,000156	0,000294
30	0,000169	0,000227	0,000239	0,000166	0,000157	0,000306
31	0,000205	0,000233	0,000249	0,000170	0,000160	0,000325
32	0,000227	0,000232	0,000252	0,000168	0,000159	0,000336
33	0,000225	0,000236	0,000262	0,000171	0,000162	0,000356
34	0,000209	0,000241	0,000273	0,000174	0,000164	0,000380
35	0,000267	0,000387	0,000301	0,000373	0,000373	0,000403
36	0,000315	0,000398	0,000314	0,000383	0,000384	0,000433
37	0,000365	0,000410	0,000330	0,000395	0,000396	0,000467
38	0,000344	0,000416	0,000341	0,000400	0,000402	0,000499
39	0,000364	0,000429	0,000359	0,000412	0,000414	0,000543
40	0,000482	0,000434	0,000371	0,000416	0,000419	0,000583
41	0,000436	0,000454	0,000396	0,000434	0,000438	0,000651
42	0,000546	0,000452	0,000404	0,000432	0,000436	0,000697
43	0,000596	0,000472	0,000432	0,000450	0,000455	0,000791
44	0,000646	0,000483	0,000455	0,000460	0,000466	0,000889
45	0,000757	0,000787	0,000751	0,001043	0,000726	0,000467
46	0,000805	0,000815	0,000817	0,001090	0,000751	0,000492
47	0,000916	0,000851	0,000901	0,001150	0,000782	0,000525
48	0,001009	0,000875	0,000986	0,001196	0,000802	0,000551
49	0,001192	0,000917	0,001111	0,001269	0,000839	0,000593
50	0,001276	0,000940	0,001238	0,001319	0,000858	0,000624
51	0,001370	0,001002	0,001458	0,001427	0,000912	0,000686
52	0,001694	0,001021	0,001676	0,001479	0,000926	0,000723
53	0,001775	0,001084	0,002066	0,001598	0,000979	0,000797
54	0,001900	0,001118	0,002590	0,001683	0,001006	0,000858
55	0,002133	0,001911	0,000975	0,001881	0,001919	0,000959
56	0,002350	0,002060	0,001076	0,002018	0,002071	0,001060
57	0,002675	0,002270	0,001220	0,002213	0,002287	0,001206
58	0,002897	0,002465	0,001375	0,002389	0,002490	0,001362
59	0,002996	0,002780	0,001625	0,002674	0,002818	0,001617
60	0,003356	0,003093	0,001926	0,002948	0,003148	0,001927
61	0,003495	0,003631	0,002472	0,003419	0,003715	0,002498
62	0,003726	0,004198	0,003274	0,003889	0,004327	0,003361
63	0,004240	0,005147	0,005086	0,004659	0,005363	0,005412
64	0,004447	0,006580	0,011214	0,005748	0,006976	0,013649

Fuente: Elaboración propia

Figura 5.2

Proyección de las tasas de mortalidad reales y por modelo para las mujeres, 2030



Fuente: Elaboración propia

Finalmente, se muestran en la tabla 5.2 y en la figura 5.2 las tasas de mortalidad proyectadas para las mujeres. Al igual que para los hombres, la primera columna de la tabla 5.2 corresponde a las tasas de mortalidad reales proyectadas al año 2030, mientras que el resto corresponde a los últimos cinco GLM mostrados en los apartados anteriores. En la figura 5.2 se ve cómo las tasas de mortalidad reales son menores a algunos modelos, entre los 11 y los 40 años de edad aproximadamente. Una vez más, entre los 50 y los 60 años, las tasas reales se sitúan por encima de las de los modelos y, a partir de ahí, vuelven a verse superadas por estos.

Tabla 5.3*Diferencias entre las tasas de mortalidad reales y de los modelos para los hombres entre 11 y 64 años*

Edades	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
11	-0,000002	0,000223	0,000003	-0,000010	-0,000004
12	0,000024	0,000252	0,000029	0,000016	0,000022
13	-0,000004	0,000234	0,000002	-0,000012	-0,000006
14	-0,000007	0,000224	-0,000002	-0,000015	-0,000009
15	0,000001	0,000256	0,000151	0,000090	-0,000012
16	-0,000009	0,000427	0,000121	0,000058	0,000016
17	-0,000028	0,000409	0,000102	0,000038	-0,000003
18	0,000071	0,000029	0,000045	0,000351	0,000056
19	0,000013	-0,000028	-0,000011	0,000297	-0,000002
20	-0,000035	-0,000076	-0,000058	0,000257	-0,000050
21	-0,000004	-0,000045	-0,000026	0,000294	-0,000019
22	-0,000004	-0,000046	-0,000025	0,000302	-0,000020
23	-0,000067	-0,000110	-0,000087	0,000243	-0,000083
24	0,000026	-0,000016	0,000008	0,000339	0,000010
25	-0,000011	-0,000071	0,000263	0,000009	-0,000006
26	0,000097	0,000037	0,000392	0,000119	0,000103
27	0,000061	0,000000	0,000373	0,000082	0,000066
28	0,000069	0,000009	0,000401	0,000091	0,000075
29	0,000055	-0,000007	0,000414	0,000078	0,000061
30	-0,000029	-0,000091	0,000354	-0,000005	-0,000023
31	-0,000048	-0,000111	0,000367	-0,000023	-0,000042
32	-0,000074	-0,000137	0,000361	-0,000050	-0,000069
33	-0,000030	-0,000094	0,000441	-0,000005	-0,000024
34	-0,000056	-0,000121	0,000456	-0,000029	-0,000050
35	0,000255	0,000197	0,000259	0,000242	0,000273
36	0,000184	0,000127	0,000201	0,000171	0,000202
37	0,000131	0,000074	0,000162	0,000118	0,000149
38	0,000120	0,000064	0,000167	0,000108	0,000139
39	-0,000024	-0,000080	0,000041	-0,000036	-0,000004
40	-0,000002	-0,000057	0,000082	-0,000014	0,000018
41	-0,000056	-0,000111	0,000051	-0,000068	-0,000035
42	-0,000090	-0,000144	0,000039	-0,000102	-0,000069
43	-0,000213	-0,000267	-0,000054	-0,000224	-0,000192
44	-0,000270	-0,000323	-0,000083	-0,000281	-0,000249
45	0,000646	0,000322	-0,000093	-0,000002	0,000627
46	0,000688	0,000351	-0,000066	-0,000006	0,000665
47	0,000508	0,000154	-0,000266	-0,000243	0,000480
48	0,000317	-0,000052	-0,000468	-0,000490	0,000285
49	0,000240	-0,000149	-0,000560	-0,000636	0,000201
50	0,000065	-0,000342	-0,000739	-0,000881	0,000019
51	-0,000149	-0,000581	-0,000962	-0,001188	-0,000203
52	-0,000457	-0,000909	-0,001259	-0,001583	-0,000522
53	-0,000873	-0,001349	-0,001656	-0,002103	-0,000949
54	-0,000925	-0,001432	-0,001678	-0,002285	-0,001016
55	-0,000624	-0,000965	-0,002212	-0,000533	-0,000561
56	-0,000689	-0,001024	-0,002348	-0,000570	-0,000624
57	-0,000731	-0,001053	-0,002465	-0,000575	-0,000666
58	-0,000868	-0,001160	-0,002646	-0,000663	-0,000802
59	-0,000755	-0,000994	-0,002558	-0,000477	-0,000692
60	-0,000688	-0,000828	-0,002433	-0,000303	-0,000629
61	-0,000602	-0,000557	-0,002111	-0,000043	-0,000553
62	-0,000045	0,000353	-0,000793	0,000783	-0,000020
63	0,001063	0,002228	0,002712	0,002410	0,001038
64	0,003188	0,006257	0,015910	0,005626	0,003047

Fuente: Elaboración propia

Tabla 5.4*Diferencias entre las tasas de mortalidad reales y de los modelos para las mujeres entre 11 y 64 años*

Edades	Modelo 6	Modelo 7	Modelo 8	Modelo 9	Modelo 10
11	0,000003	0,000098	0,000002	0,000155	0,000075
12	0,000009	0,000106	0,000008	0,000162	0,000081
13	0,000018	0,000123	0,000017	0,000181	0,000093
14	-0,000021	0,000086	-0,000022	0,000144	0,000054
15	-0,000042	-0,000010	-0,000008	0,000160	0,000108
16	-0,000041	-0,000007	-0,000006	0,000167	0,000118
17	-0,000053	-0,000018	-0,000018	0,000161	0,000116
18	0,000102	0,000045	0,000026	0,000016	0,000128
19	0,000109	0,000052	0,000031	0,000022	0,000142
20	0,000107	0,000051	0,000029	0,000019	0,000147
21	0,000116	0,000061	0,000038	0,000028	0,000163
22	0,000105	0,000052	0,000027	0,000018	0,000158
23	0,000088	0,000036	0,000010	0,000000	0,000151
24	0,000076	0,000025	-0,000003	-0,000012	0,000148
25	0,000081	0,000075	0,000025	0,000016	0,000115
26	0,000068	0,000065	0,000012	0,000003	0,000108
27	0,000067	0,000068	0,000010	0,000001	0,000116
28	0,000068	0,000071	0,000010	0,000001	0,000125
29	0,000079	0,000087	0,000019	0,000010	0,000148
30	0,000059	0,000070	-0,000002	-0,000012	0,000138
31	0,000028	0,000044	-0,000035	-0,000045	0,000120
32	0,000004	0,000025	-0,000059	-0,000068	0,000109
33	0,000011	0,000037	-0,000054	-0,000064	0,000131
34	0,000032	0,000064	-0,000035	-0,000044	0,000171
35	0,000120	0,000034	0,000107	0,000107	0,000137
36	0,000083	0,000000	0,000069	0,000069	0,000118
37	0,000045	-0,000035	0,000030	0,000031	0,000103
38	0,000072	-0,000003	0,000056	0,000058	0,000155
39	0,000066	-0,000005	0,000048	0,000051	0,000179
40	-0,000048	-0,000111	-0,000066	-0,000063	0,000101
41	0,000018	-0,000040	-0,000002	0,000002	0,000215
42	-0,000094	-0,000142	-0,000114	-0,000110	0,000151
43	-0,000124	-0,000163	-0,000146	-0,000141	0,000195
44	-0,000163	-0,000191	-0,000186	-0,000181	0,000242
45	0,000030	-0,000006	0,000286	-0,000031	-0,000290
46	0,000010	0,000012	0,000285	-0,000055	-0,000313
47	-0,000065	-0,000015	0,000234	-0,000134	-0,000391
48	-0,000134	-0,000023	0,000187	-0,000206	-0,000457
49	-0,000274	-0,000081	0,000078	-0,000353	-0,000599
50	-0,000336	-0,000038	0,000043	-0,000418	-0,000652
51	-0,000368	0,000088	0,000057	-0,000458	-0,000684
52	-0,000672	-0,000018	-0,000215	-0,000768	-0,000971
53	-0,000692	0,000290	-0,000178	-0,000797	-0,000978
54	-0,000782	0,000690	-0,000218	-0,000894	-0,001042
55	-0,000221	-0,001158	-0,000252	-0,000214	-0,001174
56	-0,000291	-0,001275	-0,000332	-0,000279	-0,001291
57	-0,000406	-0,001455	-0,000463	-0,000388	-0,001470
58	-0,000432	-0,001523	-0,000508	-0,000407	-0,001536
59	-0,000216	-0,001371	-0,000322	-0,000179	-0,001379
60	-0,000263	-0,001430	-0,000408	-0,000208	-0,001429
61	0,000136	-0,001023	-0,000077	0,000220	-0,000997
62	0,000471	-0,000453	0,000163	0,000600	-0,000365
63	0,000907	0,000847	0,000419	0,001123	0,001172
64	0,002133	0,006767	0,001301	0,002529	0,009202

Fuente: Elaboración propia

Con el propósito de analizar en mayor medida las diferencias en las proyecciones entre las tasas de mortalidad de cada modelo con respecto a las tasas de mortalidad originales, se han elaborado dos tablas, la tabla 5.3 y la tabla 5.4.

La tabla 5.3 representa las diferencias entre las tasas de mortalidad de los modelos 1 a 5 y las reales, entre las edades 11 y 64 años. Como se puede apreciar, los resultados son coherentes con los comentarios presentados junto con la figura 5.1, es decir, las tasas de los modelos son generalmente superiores a las reales entre los 11 y los 40 años, pasan a ser inferiores entre los 50 y los 60, y finalmente, vuelven a crecer, tomando unos valores mucho mayores en los últimos años de las series. La mayor diferencia en las tasas es la del modelo 3 a los 64 años, mientras que la menor, es la que corresponde al modelo 2, a los 27 años.

Para finalizar, la tabla 5.4 presenta las diferencias entre las tasas de mortalidad de las mujeres, o sea entre los modelos 6 a 10, y las reales. El comportamiento es muy similar al de los hombres, siendo incluso más favorable, ya que las tasas son en su mayoría superiores entre el principio de la serie y los 40 años de edad. La mayor diferencia se encuentra en el modelo 10, a la edad de 64 años, y la menor es la del modelo 9 a los 23 años.

6. CONCLUSIONES

El objetivo principal del presente análisis ha sido comprender cómo los diferentes estilos de vida y entornos sociales pueden influir en la salud y qué efectos tienen en el medio plazo, distinguiendo entre hombres y mujeres. Para ello, se ha investigado acerca de los factores más relevantes que afectan a las tasas de mortalidad en España y se ha obtenido que los principales son: el tabaquismo, el alcoholismo, el consumo de drogas, el ejercicio físico, el Índice de Masa Corporal, las horas de sueño, el nivel de renta y la edad. Se ha conseguido abordar esa meta principal gracias al uso de los Modelos Lineales Generalizados (GLM), con los que se han creado futuros escenarios hipotéticos, y al uso del modelo de Lee-Carter, con el que se han proyectado las tasas de mortalidad de dichos escenarios hipotéticos, para así poder visualizar la influencia de las variables sobre la mortalidad.

Los resultados obtenidos muestran las tasas de mortalidad de los diferentes escenarios para el año 2020 y las tasas proyectadas al año 2030. Para el año 2020 se ha obtenido que el modelo que proporciona, en promedio, unas tasas de mortalidad más bajas en los primeros años de la serie (es decir, entre los 11 y los 37 años) sería el de un hombre que consuma tabaco diariamente, haya consumido cannabis en el último año, se ejercite entre uno y dos días semanales, tenga sobrepeso y duerma el tiempo diario recomendado, o el de la mujer que fume de forma ocasional, haya consumido alcohol y cannabis en el último año, se ejercite entre una y dos veces por semana y tenga sobrepeso. Estos resultados pueden estar sesgados por la falta de datos para los primeros años de la primera mitad de la serie, pues no existen registros del consumo de tabaco, alcohol o drogas para personas menores a los 15 años y, por ello, pueden no tener el peso suficiente sobre las tasas de mortalidad. En cuanto a los últimos años de la serie (de 38 a 64 años), los modelos que implican menores tasas de mortalidad serían el de un hombre del que únicamente se sepa que ha consumido alcohol en el último año y se vea afectado por los evidentes efectos de la edad, o el de una mujer que únicamente se conozca que se ejercita entre tres y cuatro días por semana y esté influida por la edad. En este caso, dado que se desconoce el estado físico tanto del hombre, como de la mujer, se tiene que la influencia de los factores mencionados es positiva para la salud, o al menos, en mayor medida que en el caso del resto de modelos.

Por otro lado, se tiene que los resultados de las proyecciones de las tasas de mortalidad con el modelo de Lee-Carter al año 2030 indican que las menores tasas se vinculan a los

mismos individuos que en el año 2020. Sin embargo, si se analiza la serie completa, se tiene que, en promedio, el hombre que tendrá menores tasas de mortalidad será aquel del que únicamente se sepa que ha consumido alcohol en los últimos doce meses, y al que se le atribuyen tasas más altas es aquel que haga ejercicio uno o dos días semanales, tenga sobrepeso, duerma correctamente, fume a diario y haya consumido cannabis alguna vez en el último año, o bien aquel que fume ocasionalmente, haya consumido hipnosedantes, realice ejercicio cinco o seis días por semana a pesar de tener un Índice de Masa Corporal normal y duerma correctamente. Gracias a la proyección de las tasas de mortalidad, se puede ver cómo las variables dañinas que se han tenido en cuenta en este análisis tienen efectos negativos en el medio plazo, es decir, que cuando se estaba analizando la serie en el año 2020, había un modelo que proporcionaba tasas de mortalidad muy bajas, a pesar de tener en cuenta variables poco saludables (como el consumo de determinadas drogas o tabaco), pero que con el paso del tiempo, esas variables han afectado de forma muy negativa en su salud, tanto que, ese modelo que devolvía unas tasas de mortalidad tan bajas se ha convertido en el más mortal. En cuanto a las mujeres en el año 2030, en promedio, la más longeva será aquella que realice ejercicio entre tres y cuatro días por semana, mientras que aquella con las tasas de mortalidad más elevadas será la que fume de manera ocasional, haya consumido alcohol y cannabis en el último año, tenga sobrepeso y se ejercite uno o dos días por semana. Una vez más, a través de las proyecciones del modelo Lee-Carter, se ha podido observar cómo determinados factores perjudiciales para la salud han tenido efectos mucho mayores con el paso del tiempo, como es el caso del tabaco, las drogas o el alcohol.

También se han comparado las proyecciones de los diferentes escenarios de hombres y mujeres con las tasas de mortalidad reales llevadas al año 2030. En promedio, los modelos de las mujeres se ajustan más que los de los hombres a las tasas reales.

En conclusión, este estudio ha proporcionado un minucioso análisis del efecto de los hábitos de vida sobre las tasas de mortalidad, de modo que se haya comprendido más profundamente la importancia de llevar una vida más saludable. Asimismo, se han sentado las bases para posibles trabajos futuros, en los que se podría profundizar en la obtención de más factores que influyan en los fallecimientos o, incluso, segmentar el estudio por Comunidades Autónomas, de forma que se conciencie a población más concreta para mejorar sus hábitos de vida y promover el bienestar común.

BIBLIOGRAFÍA

- Debón Aucejo, A., Martínez Ruíz, F., Montes Suay, F., & Moshuk, M. (2015). En *Proyección de los indicadores de mortalidad para España* (págs. 291-321). Instituto Nacional de Estadística.
- Gisbert, F. J., & Pallejà, R. P. (2006). *Esperanza de vida en España a lo largo del siglo XX*. Obtenido de Fundación BBVA: https://www.fbbva.es/wp-content/uploads/2017/05/dat/DT_2006_11.pdf
- INE. (2018). *Movimiento Natural de la Población. Indicadores Demográficos Básicos*. Madrid: Instituto Nacional de Estadística. Obtenido de https://www.ine.es/prensa/mnp_2018_p.pdf
- INE. (2023). *Indicadores de Calidad de Vida*. Madrid: Instituto Nacional de Estadística.
- INE. (2023). *Tablas de mortalidad. Metodología*. Obtenido de <https://www.ine.es/metodologia/t20/t2020319a.pdf>
- Ministerio de Sanidad. (2018). *Encuesta Nacional de Salud. España 2017*. Obtenido de https://www.sanidad.gob.es/estadEstudios/estadisticas/encuestaNacional/encuestaNac2017/ENSE2017_notatecnica.pdf
- Ministerio de Sanidad, Servicios Sociales e Igualdad. (2016). *Informe Anual del Sistema Nacional de Salud 2016*.
- National Heart, Lung and Blood Institute. (2022). *National Heart, Lung and Blood Institute. SLEEP DERIVATION AND DEFICIENCY. How Much Sleep Is Enough*. Obtenido de <https://www.nhlbi.nih.gov/health/sleep-deprivation/how-much-sleep>
- Observatorio Español de las Drogas y las Adicciones. (2023). *Informe 2023. Alcohol, tabaco y drogas ilegales en España*. Madrid: Ministerio de Sanidad. Delegación del Gobierno para el Plan Nacional sobre Drogas.
- OECD. (2023). *Health at a Glance 2023: OECD Indicators*. París: OECD Publishing.
- Organización Mundial de la Salud. (2019). *Plan de acción mundial sobre actividad física 2018-2030: personas más activas para un mundo más sano*. Obtenido de Organización Mundial de la Salud: <https://iris.who.int/handle/10665/327897>
- Organización Mundial de la Salud. (2023). *WHO acceleration plan to stop obesity*. Obtenido de

<https://iris.who.int/bitstream/handle/10665/370281/9789240075634-eng.pdf?sequence=1>

Organización Mundial de la Salud. (2024). *Organización Mundial de la Salud. Obesidad y sobrepeso*. Obtenido de <https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight>

Rodríguez-Pardo del Castillo, J. M., & López Farré, A. (2017). *LONGEVIDAD y ENVEJECIMIENTO en el tercer milenio: nuevas perspectivas*. Madrid: Fundación Mapfre.

ANEXO

```
options(scipen = 999)
```

```
library("readxl")
```

```
library("ggplot2")
```

```
library("demography")
```

```
library("corrplot")
```

```
library("PerformanceAnalytics")
```

```
library("fitdistrplus")
```

```
library("data.table")
```

```
#####GRÁFICOS Y CORRELACIONES#####
```

```
directorio<-"C:/Users/ricar/OneDrive/Escritorio/Máster/TFM/Base de Datos/"
```

```
setwd(directorio)
```

```
#####GRÁFICO DE ESPERANZA DE VIDA#####
```

```
Esperanza_vida <- read_excel(path = "Datos/Esperanza de vida.xlsx", sheet = "D1T1", range =  
"A3:E35", col_names = TRUE)
```

```
Esperanza_vida <- subset(Esperanza_vida, select = -c(2, 4))
```

```
Esperanza_vida <- ts(Esperanza_vida)
```

```
#Extraemos años y evolución por separado
```

```
años_esp_vida <- as.vector(Esperanza_vida[,1])
```

```
evol_esp_vida_h <- as.vector(Esperanza_vida[,2])
```

```
evol_esp_vida_m <- as.vector(Esperanza_vida[,3])
```

```
#Creamos un data frame combinando años y evolución
```

```
datos_esperanza_vida <- data.frame(años = años_esp_vida, evolucion_h = evol_esp_vida_h,  
evolucion_m = evol_esp_vida_m)
```

```

#Graficamos la serie de tiempo

ggplot(data = datos_esperanza_vida, aes(x = años))+
  geom_line(aes(y=evolucion_h,color="Hombres"))+
  geom_line(aes(y=evolucion_m,color="Mujeres")) +
  scale_color_manual(values=c("Hombres"="red","Mujeres"="blue")) +
  labs(title = "Evolución de la Esperanza de Vida 1991-2022", x = "Años", y = "Esperanza de
vida") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

```

```
#####GRÁFICO DE TABAQUISMO#####
```

```

Tabaquismo <- read_excel(path="Datos/tabaco.xlsx",
sheet="D7T1",range="A6:E40",col_names = FALSE)

Tabaquismo <- subset(Tabaquismo, select = -c(2, 4))

grupos_tabaco <- unlist(Tabaquismo[1:8,1])
grupos_tabaco <- as.vector(grupos_tabaco)

#Fumadores diarios

fumador_diario_h <- unlist(Tabaquismo[1:8,2])
fumador_diario_h <- as.numeric(fumador_diario_h)

fumador_diario_m <- unlist(Tabaquismo[1:8,3])
fumador_diario_m <- as.numeric(fumador_diario_m)

df_diario <- data.frame(grupos_tabaco, fumador_diario_h, fumador_diario_m)

ggplot(data = df_diario, aes(x = grupos_tabaco)) +
  geom_bar(aes(y = fumador_diario_h, fill = "Hombres"), stat = "identity", position =
position_dodge(width = 0.8)) + geom_bar(aes(y = fumador_diario_m, fill = "Mujeres"), stat =
"identity", position = position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("Hombres" = "red", "Mujeres" = "blue")) +
  labs(title = "% de fumadores diarios por sexo y grupos de edad, 2020",
x = "Grupos de edad",

```

```

y = "Fumadores diarios (%)" +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))
#Fumadores ocasionales
fumador_ocasional_h <- unlist(Tabaquismo[10:17,2])
fumador_ocasional_h <- as.numeric(fumador_ocasional_h)
fumador_ocasional_m <- unlist(Tabaquismo[10:17,3])
fumador_ocasional_m <- as.numeric(fumador_ocasional_m)
df_ocasional <- data.frame(grupos_tabaco, fumador_ocasional_h, fumador_ocasional_m)
ggplot(data = df_ocasional, aes(x = grupos_tabaco)) +
  geom_bar(aes(y = fumador_ocasional_h, fill = "Hombres"), stat = "identity", position =
position_dodge(width = 0.8)) +
  geom_bar(aes(y = fumador_ocasional_m, fill = "Mujeres"), stat = "identity", position =
position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("Hombres" = "red", "Mujeres" = "blue")) +
  labs(title = "% de fumadores ocasionales por sexo y grupos de edad, 2020",
x = "Grupos de edad",
y = "Fumadores ocasionales (%)") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))
#Exfumadores
exfumador_h <- unlist(Tabaquismo[19:26,2])
exfumador_h <- as.numeric(exfumador_h)
exfumador_m <- unlist(Tabaquismo[19:26,3])
exfumador_m <- as.numeric(exfumador_m)
df_exfumador <- data.frame(grupos_tabaco, exfumador_h, exfumador_m)
ggplot(data = df_exfumador, aes(x = grupos_tabaco)) +
  geom_bar(aes(y = exfumador_h, fill = "Hombres"), stat = "identity", position =
position_dodge(width = 0.8)) +

```



```

geom_bar(aes(y = exfumador_m, fill = "Mujeres"), stat = "identity", position =
position_dodge(width = 0.8)) +
scale_fill_manual(values = c("Hombres" = "red", "Mujeres" = "blue")) +
labs(title = "% de exfumadores por sexo y grupos de edad, 2020",
x = "Grupos de edad",
y = "Exfumadores (%)") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))

#No fumadores
no_fumador_h <- unlist(Tabaquismo[28:35,2])
no_fumador_h <- as.numeric(no_fumador_h)
no_fumador_m <- unlist(Tabaquismo[28:35,3])
no_fumador_m <- as.numeric(no_fumador_m)
df_no_fumador <- data.frame(grupos_tabaco, no_fumador_h, no_fumador_m)
ggplot(data = df_no_fumador, aes(x = grupos_tabaco)) +
geom_bar(aes(y = no_fumador_m, fill = "Mujeres"), stat = "identity", position =
position_dodge(width = 0.8)) +
geom_bar(aes(y = no_fumador_h, fill = "Hombres"), stat = "identity", position =
position_dodge(width = 0.8)) +
scale_fill_manual(values = c("Hombres" = "red", "Mujeres" = "blue")) +
labs(title = "% de no fumadores por sexo y grupos de edad, 2020",
x = "Grupos de edad",
y = "No fumadores (%)") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))

#####GRÁFICO DE ALCOHOLISMO#####
Alcoholismo <- read_excel(path="Datos/alcohol.xlsx",
sheet="D7T3",range="A6:E13",col_names = FALSE)

```

```

Alcoholismo <- subset(Alcoholismo, select = -c(2, 4))
grupos_alcohol <- unlist(Alcoholismo[1:8,1])
grupos_alcohol <- as.vector(grupos_alcohol)
#Consumo de alcohol en los últimos 12 meses
alcohol_h <- unlist(Alcoholismo[1:8,2])
alcohol_h <- as.numeric(alcohol_h)
alcohol_m <- unlist(Alcoholismo[1:8,3])
alcohol_m <- as.numeric(alcohol_m)
df_alcohol <- data.frame(grupos_alcohol, alcohol_h, alcohol_m)
ggplot(data = df_alcohol, aes(x = grupos_alcohol)) +
  geom_bar(aes(y = alcohol_h, fill = "Hombres"), stat = "identity", position =
position_dodge(width = 0.8)) +
  geom_bar(aes(y = alcohol_m, fill = "Mujeres"), stat = "identity", position =
position_dodge(width = 0.8)) +
  scale_fill_manual(values = c("Hombres" = "red", "Mujeres" = "blue")) +
  labs(title = "% de consumidores de alcohol en los últimos 12 meses por sexo y grupos de edad,
2020",
  x = "Grupos de edad",
  y = "Consumidores de alcohol (%)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

```

#####GRÁFICO DE DROGAS#####

```

Drogas <- read_excel(path="Datos/drogas.xlsx", sheet="OEDA",range="A3:Y20",col_names =
FALSE)
Drogas <- Drogas[c(-2,-3),]
años_drogas <- c(1999,2001,2003,2005,2007,2009,2011,2013,2015,2018,2020,2022)
grupos_drogas <- unlist(Drogas[2:16,1])
grupos_drogas <- as.vector(grupos_drogas)

```

```

####Consumo de drogas en los últimos 12 meses

grupo_15_34 <- seq(from=2, to=24, by=2)
drogas_15_34 <- Drogas[2:16,grupo_15_34]
grupo_35_64 <- seq(from=3, to=25, by=2)
drogas_35_64 <- Drogas[2:16,grupo_35_64]

#Cannabis 15-34
cannabis_15_34 <- t(matrix(drogas_15_34[5,], nrow = 1))
plot(años_drogas,cannabis_15_34,type="l",col="green",xlab="Años",ylab="% de consumo de
cannabis", main="Consumo de cannabis en los últimos 12 meses para población de 15-34 años,
España, 1999-2022")

#Cannabis 35-64
cannabis_35_64 <- t(matrix(drogas_35_64[5,], nrow = 1))
plot(años_drogas,cannabis_35_64,type="l",col="orange",xlab="Años",ylab="% de consumo de
cannabis", main="Consumo de cannabis en los últimos 12 meses para población de 35-64 años,
España, 1999-2022")

#Cocaína 15-34
cocaina_15_34 <- t(matrix(drogas_15_34[9,], nrow = 1))
plot(años_drogas,cocaina_15_34,type="l",col="green",xlab="Años",ylab="% de consumo de
cocaína", main="Consumo de cocaína en los últimos 12 meses para población de 15-34 años,
España, 1999-2022")

#Cocaína 35-64
cocaina_35_64 <- t(matrix(drogas_35_64[9,], nrow = 1))
plot(años_drogas,cocaina_35_64,type="l",col="orange",xlab="Años",ylab="% de consumo de
cocaína", main="Consumo de cocaína en los últimos 12 meses para población de 35-64 años,
España, 1999-2022")

#Hipnosedantes 15-34
hipnosedantes_15_34 <- t(matrix(drogas_15_34[1,], nrow = 1))
plot(años_drogas,hipnosedantes_15_34,type="l",col="green",xlab="Años",ylab="%
de consumo de hipnosedantes", main="Consumo de hipnosedantes en los últimos 12 meses para
población de 15-34 años, España, 1999-2022")

```

```
#Hipnosedantes 35-64
```

```
hipnosedantes_35_64 <- t(matrix(drogas_35_64[1,], nrow = 1))
```

```
plot(años_drogas,hipnosedantes_35_64,type="l",col="orange",xlab="Años",ylab="% de  
consumo de hipnosedantes", main="Consumo de hipnosedantes en los últimos 12 meses para  
población de 35-64 años, España, 1999-2022")
```

```
#####GRÁFICO EVOLUCIÓN EJERCICIO FÍSICO#####
```

```
Ejercicio <- read_excel(path="Datos/ejercicio_fisico.xlsx",  
sheet="sexo",range="A3:F7",col_names = FALSE)
```

```
años_ejercicio <- as.vector(unlist(Ejercicio[1,2:6]))
```

```
ejercicio_total <- as.numeric(Ejercicio[3,2:6])
```

```
ejercicio_h <- as.numeric(Ejercicio[4,2:6])
```

```
ejercicio_m <- as.numeric(Ejercicio[5,2:6])
```

```
df_ejercicio <- data.frame(años_ejercicio, ejercicio_total, ejercicio_h, ejercicio_m)
```

```
# Graficamos la serie de tiempo
```

```
ggplot(data = df_ejercicio, aes(x = años_ejercicio))+
```

```
geom_line(aes(y=ejercicio_total,color="Total", group=1))+
```

```
geom_line(aes(y=ejercicio_h,color="Hombres", group=1))+
```

```
geom_line(aes(y=ejercicio_m,color="Mujeres", group=1)) +
```

```
scale_color_manual(values=c("Total"="green","Hombres"="red","Mujeres"="blue")) +
```

```
labs(title = "Evolución de la práctica de ejercicio físico 2011-2022", x = "Años", y = "% de  
población que se ejercita regularmente") +
```

```
theme_minimal() +
```

```
theme(plot.title = element_text(hjust = 0.5))
```

```
#####GRÁFICO DÍAS DE EJERCICIO FÍSICO#####
```

```
Días_Ejercicio <- read_excel(path="Datos/ejercicio_fisico por grupos de edad.xlsx",  
sheet="D6T5",range="A3:K13",col_names = FALSE)
```

```
Días_Ejercicio <- subset(Días_Ejercicio,select=-c(2,4,6,8,10))
```

```

días_ejercicio <- as.vector(unlist(Días_Ejercicio[1,2:6]))

####Hombres

h_15_24 <- as.numeric(Días_Ejercicio[3,2:6])
h_25_44 <- as.numeric(Días_Ejercicio[4,2:6])
h_45_64 <- as.numeric(Días_Ejercicio[5,2:6])
h_65_más <- as.numeric(Días_Ejercicio[6,2:6])

df_días_ejercicio_h <- data.frame(días_ejercicio, h_15_24, h_25_44, h_45_64, h_65_más)

#Graficamos para hombres

ggplot(data = df_días_ejercicio_h, aes(x = días_ejercicio))+
  geom_line(aes(y=h_15_24,color="Entre 15 y 24 años", group=1))+
  geom_line(aes(y=h_25_44,color="Entre 25 y 44 años", group=1))+
  geom_line(aes(y=h_45_64,color="Entre 45 y 64 años", group=1)) +
  geom_line(aes(y=h_65_más,color="65 o más años", group=1)) +
  scale_color_manual(values=c("Entre 15 y 24 años"="green","Entre 25 y 44 años"="red","Entre
45 y 64 años"="blue","65 o más años"="purple")) +
  labs(title = "Número de días de ejercicio físico semanal, para hombres", x = "Días a la semana",
y = "% de población por grupos de edad") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

####Mujeres

m_15_24 <- as.numeric(Días_Ejercicio[8,2:6])
m_25_44 <- as.numeric(Días_Ejercicio[9,2:6])
m_45_64 <- as.numeric(Días_Ejercicio[10,2:6])
m_65_más <- as.numeric(Días_Ejercicio[11,2:6])

df_días_ejercicio_m <- data.frame(días_ejercicio, m_15_24, m_25_44, m_45_64, m_65_más)

#Graficamos para mujeres

ggplot(data = df_días_ejercicio_m, aes(x = días_ejercicio))+
  geom_line(aes(y=m_15_24,color="Entre 15 y 24 años", group=1))+
  geom_line(aes(y=m_25_44,color="Entre 25 y 44 años", group=1))+

```

```

geom_line(aes(y=m_45_64,color="Entre 45 y 64 años", group=1)) +
geom_line(aes(y=m_65_más,color="65 o más años", group=1)) +
scale_color_manual(values=c("Entre 15 y 24 años"="green","Entre 25 y 44 años"="red","Entre
45 y 64 años"="blue","65 o más años"="purple")) +
labs(title = "Número de días de ejercicio físico semanal, para mujeres", x = "Días a la semana",
y = "% de población por grupos de edad") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))

```

#####GRÁFICO IMC#####

```

IMC <- read_excel(path="Datos/IMC.xlsx", sheet="tabla-0", range="A7:F37",
col_names=TRUE)

```

```

IMC <- IMC[c(-1:-10),-2]

```

```

años_IMC <- unlist(IMC[3:10,1])

```

```

años_IMC <- as.vector(años_IMC)

```

###Hombres

```

IMC_h <- IMC[3:10,2:5]

```

```

Insuficiente_h <- as.numeric(unlist(IMC_h[,1]))

```

```

Normopeso_h <- as.numeric(unlist(IMC_h[,2]))

```

```

Sobrepeso_h <- as.numeric(unlist(IMC_h[,3]))

```

```

Obesidad_h <- as.numeric(unlist(IMC_h[,4]))

```

```

df_IMC_h <- data.frame(años_IMC, Insuficiente_h, Normopeso_h, Sobrepeso_h, Obesidad_h)

```

```

ggplot(data = df_IMC_h, aes(x = años_IMC)) +

```

```

geom_line(aes(y = Insuficiente_h, color = "Insuficiente",group=1)) +

```

```

geom_line(aes(y = Normopeso_h, color = "Normopeso",group=1)) +

```

```

geom_line(aes(y = Sobrepeso_h, color = "Sobrepeso",group=1)) +

```

```

geom_line(aes(y = Obesidad_h, color = "Obesidad",group=1)) +

```

```

scale_color_manual(values = c("Insuficiente" = "red", "Normopeso" = "blue", "Sobrepeso" =
"green", "Obesidad" = "purple")) +

```

```

labs(title = "Índice de Masa Corporal por grupos de edad para hombres, 2020", x = "Edades", y
= "Índice de Masa Corporal") +

theme_minimal() +

theme(plot.title = element_text(hjust = 0.5))

###Mujeres

IMC_m <- IMC[13:20,2:5]

Insuficiente_m <- as.numeric(unlist(IMC_m[,1]))

Normopeso_m <- as.numeric(unlist(IMC_m[,2]))

Sobrepeso_m <- as.numeric(unlist(IMC_m[,3]))

Obesidad_m <- as.numeric(unlist(IMC_m[,4]))

df_IMC_m <- data.frame(años_IMC, Insuficiente_m, Normopeso_m, Sobrepeso_m, Obesidad_m)

ggplot(data = df_IMC_m, aes(x = años_IMC)) +

  geom_line(aes(y = Insuficiente_m, color = "Insuficiente", group=1)) +

  geom_line(aes(y = Normopeso_m, color = "Normopeso", group=1)) +

  geom_line(aes(y = Sobrepeso_m, color = "Sobrepeso", group=1)) +

  geom_line(aes(y = Obesidad_m, color = "Obesidad", group=1)) +

  scale_color_manual(values = c("Insuficiente" = "red", "Normopeso" = "blue", "Sobrepeso" =
"green", "Obesidad" = "purple")) +

  labs(title = "Índice de Masa Corporal por grupos de edad para mujeres, 2020", x = "Edades", y
= "Índice de Masa Corporal") +

  theme_minimal() +

  theme(plot.title = element_text(hjust = 0.5))

#####GRÁFICO HORAS DE SUEÑO#####

Sueño <- read_excel(path="Datos/horas de sueño.xlsx", sheet="tabla-0_sueño",
range="A7:F35", col_names=TRUE)

Sueño <- Sueño[c(-1:-10),-2]

años_Sueño <- unlist(Sueño[3:9,1])

años_Sueño <- as.vector(años_Sueño)

```

```
####Hombres
```

```
Sueño_h <- Sueño[3:9,2:5]
```

```
Menos_6_h <- as.numeric(unlist(Sueño_h[,1]))
```

```
Entre_6_7_h <- as.numeric(unlist(Sueño_h[,2]))
```

```
Entre_8_9_h <- as.numeric(unlist(Sueño_h[,3]))
```

```
Mas_10_h <- as.numeric(unlist(Sueño_h[,4]))
```

```
df_Sueño_h <- data.frame(años_Sueño, Menos_6_h,Entre_6_7_h,Entre_8_9_h,Mas_10_h)
```

```
ggplot(data = df_Sueño_h, aes(x = años_Sueño)) +
```

```
  geom_line(aes(y = Menos_6_h, color = "Menos de 6",group=1)) +
```

```
  geom_line(aes(y = Entre_6_7_h, color = "Entre 6 y 7",group=1)) +
```

```
  geom_line(aes(y = Entre_8_9_h, color = "Entre 8 y 9",group=1)) +
```

```
  geom_line(aes(y = Mas_10_h, color = "10 o más",group=1)) +
```

```
  scale_color_manual(values = c("Menos de 6" = "red", "Entre 6 y 7" = "blue", "Entre 8 y 9" =  
"green", "10 o más" = "purple")) +
```

```
  labs(title = "% de población por horas de sueño y grupos de edad para hombres, 2020", x =  
"Edades", y = "Horas de sueño") +
```

```
  theme_minimal() +
```

```
  theme(plot.title = element_text(hjust = 0.5))
```

```
####Mujeres
```

```
Sueño_m <- Sueño[12:18,2:5]
```

```
Menos_6_m <- as.numeric(unlist(Sueño_m[,1]))
```

```
Entre_6_7_m <- as.numeric(unlist(Sueño_m[,2]))
```

```
Entre_8_9_m <- as.numeric(unlist(Sueño_m[,3]))
```

```
Mas_10_m <- as.numeric(unlist(Sueño_m[,4]))
```

```
df_Sueño_m <- data.frame(años_Sueño, Menos_6_m,Entre_6_7_m,Entre_8_9_m,Mas_10_m)
```

```
ggplot(data = df_Sueño_m, aes(x = años_Sueño)) +
```

```
  geom_line(aes(y = Menos_6_m, color = "Menos de 6",group=1)) +
```

```
  geom_line(aes(y = Entre_6_7_m, color = "Entre 6 y 7",group=1)) +
```

```
  geom_line(aes(y = Entre_8_9_m, color = "Entre 8 y 9",group=1)) +
```



```

geom_line(aes(y = Mas_10_m, color = "10 o más",group=1)) +
scale_color_manual(values = c("Menos de 6" = "red", "Entre 6 y 7" = "blue", "Entre 8 y 9" =
"green", "10 o más" = "purple")) +
labs(title = "% de población por horas de sueño y grupos de edad para mujeres, 2020", x =
"Edades", y = "Horas de sueño") +
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))

```

#####GRÁFICO RENTA#####

```

Renta <- read_excel(path="Datos/Renta.xlsx", sheet="tabla-13928", range="A8:W29",
col_names=FALSE)

```

```

Renta <- Renta[c(-1:-10),c(-2:-3)]

```

```

Renta <- Renta[c(-6:-12),-seq(2, 20, by = 2)]

```

```

años_Renta <- unlist(Renta[,1])

```

```

años_Renta <- as.vector(años_Renta)

```

```

Decil_1 <- as.numeric(unlist(Renta[,2]))

```

```

Decil_2 <- as.numeric(unlist(Renta[,3]))

```

```

Decil_3 <- as.numeric(unlist(Renta[,4]))

```

```

Decil_4 <- as.numeric(unlist(Renta[,5]))

```

```

Decil_5 <- as.numeric(unlist(Renta[,6]))

```

```

Decil_6 <- as.numeric(unlist(Renta[,7]))

```

```

Decil_7 <- as.numeric(unlist(Renta[,8]))

```

```

Decil_8 <- as.numeric(unlist(Renta[,9]))

```

```

Decil_9 <- as.numeric(unlist(Renta[,10]))

```

```

Decil_10 <- as.numeric(unlist(Renta[,11]))

```

```

df_Renta <-
data.frame(años_Renta,Decil_1,Decil_2,Decil_3,Decil_4,Decil_5,Decil_6,Decil_7,Decil_8,Decil_9,Decil_10)

```

```

df_Renta$años_Renta <- factor(df_Renta$años_Renta, levels = unique(df_Renta$años_Renta))

ggplot(data = df_Renta, aes(x = años_Renta)) +

  geom_line(aes(y = Decil_1, color = "Primer decil",group=1)) +
  geom_line(aes(y = Decil_2, color = "Segundo decil",group=1)) +
  geom_line(aes(y = Decil_3, color = "Tercer decil",group=1)) +
  geom_line(aes(y = Decil_4, color = "Cuarto decil",group=1)) +
  geom_line(aes(y = Decil_5, color = "Quinto decil",group=1)) +
  geom_line(aes(y = Decil_6, color = "Sexto decil",group=1)) +
  geom_line(aes(y = Decil_7, color = "Séptimo decil",group=1)) +
  geom_line(aes(y = Decil_8, color = "Octavo decil",group=1)) +
  geom_line(aes(y = Decil_9, color = "Noveno decil",group=1)) +
  geom_line(aes(y = Decil_10, color = "Décimo decil",group=1)) +

  scale_color_manual(values = c("Primer decil" = "red", "Segundo decil" = "blue", "Tercer decil"
= "green", "Cuarto decil"="steelblue", "Quinto decil" = "purple", "Sexto decil" = "orange",
"Séptimo decil" = "black", "Octavo decil" = "pink", "Noveno decil" = "lightblue", "Décimo decil"
= "brown"),breaks = c("Primer decil", "Segundo decil", "Tercer decil", "Cuarto decil", "Quinto
decil", "Sexto decil", "Séptimo decil", "Octavo decil", "Noveno decil", "Décimo decil")) +

  labs(title = "Salarios medios brutos por grupo de edad y decil, 2021", x = "Edades", y = "Salario
medio en euros") +

  theme_minimal() +

  theme(plot.title = element_text(hjust = 0.5))

#####CORRELACIONES MUJERES#####

Mujeres=read.table("Datos/Mujeres_qx_nuevo.txt", header = TRUE, dec = ",")

#edad_16 <- as.vector(Mujeres[16,2:63])

# Tabaco

correl_tabaco=data.frame(Mujeres$X2020, Mujeres$Fumador_diario
, Mujeres$Fumador_ocasional , Mujeres$Exfumador ,Mujeres$No_fumador)

correlacion_lineal_tabaco<-cor(correl_tabaco)

```

```

correlacion_lineal_tabaco<-corrplot(correlacion_lineal_tabaco, method="number")

tabla_corr <- chart.Correlation(correl_tabaco)

# Alcohol

correl_alcohol=data.frame(Mujeres$X2020, Mujeres$Alcohol)

correlacion_lineal_alcohol<-cor(correl_alcohol)

correlacion_lineal_alcohol<-corrplot(correlacion_lineal_alcohol, method="number")

tabla_corr <- chart.Correlation(correl_alcohol)

# Drogas

correl_drogas=data.frame(Mujeres$X2020, Mujeres$Can , Mujeres$Co , Mujeres$Hip)

correlacion_lineal_drogas<-cor(correl_drogas)

correlacion_lineal_drogas<-corrplot(correlacion_lineal_drogas, method="number")

tabla_corr <- chart.Correlation(correl_drogas)

# Ejercicio fisico

correl_ejercicio=data.frame(Mujeres$X2020, Mujeres$Ejercicio_0 , Mujeres$Ejercicio_1_2 ,
Mujeres$Ejercicio_3_4 , Mujeres$Ejercicio_5_6, Mujeres$Ejercicio_7)

correlacion_lineal_ejercicio<-cor(correl_ejercicio)

correlacion_lineal_ejercicio<-corrplot(correlacion_lineal_ejercicio, method="number")

tabla_corr <- chart.Correlation(correl_ejercicio)

# IMC

correl_IMC=data.frame(Mujeres$X2020, Mujeres$Insuficiente , Mujeres$Normopeso ,
Mujeres$Sobrepeso ,Mujeres$Obesidad)

correlacion_lineal_IMC<-cor(correl_IMC)

correlacion_lineal_IMC<-corrplot(correlacion_lineal_IMC, method="number")

tabla_corr <- chart.Correlation(correl_IMC)

# Horas de sueño

correl_sueno=data.frame(Mujeres$X2020, Mujeres$Sueno)

correlacion_lineal_sueno<-cor(correl_sueno)

correlacion_lineal_sueno<-corrplot(correlacion_lineal_sueno, method="number")

tabla_corr <- chart.Correlation(correl_sueno)

```

```

# Renta
correl_renta=data.frame(Mujeres$X2020, Mujeres$Renta)
correlacion_lineal_renta<-cor(correl_renta)
correlacion_lineal_renta<-corrplot(correlacion_lineal_renta, method="number")
tabla_corr <- chart.Correlation(correl_renta)

# Edad
correl_edad=data.frame(Mujeres$X2020, Mujeres$Edad)
correlacion_lineal_edad<-cor(correl_edad)
correlacion_lineal_edad<-corrplot(correlacion_lineal_edad, method="number")
tabla_corr <- chart.Correlation(correl_edad)

#####CORRELACIONES HOMBRES#####
Hombres=read.table("Datos/Hombres_qx_nuevo.txt", header = TRUE, dec = ",")

# Tabaco
correl_tabaco=data.frame(Hombres$X2020, Hombres$Fumador_diario ,
Hombres$Fumador_ocasional , Hombres$Exfumador ,Hombres$No_fumador)
correlacion_lineal_tabaco<-cor(correl_tabaco)
correlacion_lineal_tabaco<-corrplot(correlacion_lineal_tabaco, method="number")
tabla_corr <- chart.Correlation(correl_tabaco)

# Alcohol
correl_alcohol=data.frame(Hombres$X2020, Hombres$Alcohol)
correlacion_lineal_alcohol<-cor(correl_alcohol)
correlacion_lineal_alcohol<-corrplot(correlacion_lineal_alcohol, method="number")
tabla_corr <- chart.Correlation(correl_alcohol)

# Drogas
correl_drogas=data.frame(Hombres$X2020, Hombres$Can , Hombres$Co , Hombres$Hip)
correlacion_lineal_drogas<-cor(correl_drogas)
correlacion_lineal_drogas<-corrplot(correlacion_lineal_drogas, method="number")

```

```

tabla_corr <- chart.Correlation(correl_drogas)

# Ejercicio fisico
correl_ejercicio=data.frame(Hombres$X2020, Hombres$Ejercicio_0 , Hombres$Ejercicio_1_2 ,
Hombres$Ejercicio_3_4 , Hombres$Ejercicio_5_6, Hombres$Ejercicio_7)
correlacion_lineal_ejercicio<-cor(correl_ejercicio)
correlacion_lineal_ejercicio<-corrplot(correlacion_lineal_ejercicio, method="number")
tabla_corr <- chart.Correlation(correl_ejercicio)

# IMC
correl_IMC=data.frame(Hombres$X2020, Hombres$Insuficiente , Hombres$Normopeso ,
Hombres$Sobrepeso ,Hombres$Obesidad)
correlacion_lineal_IMC<-cor(correl_IMC)
correlacion_lineal_IMC<-corrplot(correlacion_lineal_IMC, method="number")
tabla_corr <- chart.Correlation(correl_IMC)

# Horas de sueño
correl_sueno=data.frame(Hombres$X2020, Hombres$Sueno)
correlacion_lineal_sueno<-cor(correl_sueno)
correlacion_lineal_sueno<-corrplot(correlacion_lineal_sueno, method="number")
tabla_corr <- chart.Correlation(correl_sueno)

# Renta
correl_renta=data.frame(Hombres$X2020, Hombres$Renta)
correlacion_lineal_renta<-cor(correl_renta)
correlacion_lineal_renta<-corrplot(correlacion_lineal_renta, method="number")
tabla_corr <- chart.Correlation(correl_renta)

# Edad
correl_edad=data.frame(Hombres$X2020, Hombres$Edad)
correlacion_lineal_edad<-cor(correl_edad)
correlacion_lineal_edad<-corrplot(correlacion_lineal_edad, method="number")
tabla_corr <- chart.Correlation(correl_edad)

```

```
#####MODELOS GLM#####
```

```
#####Introducimos los datos#####
```

```
Directorio="C:/Users/ricar/OneDrive/Escritorio/Máster/TFM/Base de Datos/Datos/"
```

```
setwd(Directorio)
```

```
Hombres=read.table("Hombres_qx_nuevo_11_64.txt", header = TRUE, dec = ",")
```

```
Mujeres=read.table("Mujeres_qx_nuevo_11_64.txt", header = TRUE, dec = ",")
```

```
ajuste_gamma <- fitdist(Mujeres$X2020, "gamma")
```

```
summary(ajuste_gamma)
```

```
plot(ajuste_gamma) # Parece que ajusta mejor la función de densidad y el qqplot me convence más
```

```
ajuste_gamma_h <- fitdist(Hombres$X2020, "gamma")
```

```
summary(ajuste_gamma_h)
```

```
plot(ajuste_gamma_h)
```

```
#####Una vez que conocemos la significatividad, gracias a las correlaciones, hacemos los GLMs para proyectar la mortalidad#####
```

```
#modelo_0=glm(Hombres$X2020 ~ Hombres$Fumador_diario + Hombres$Fumador_ocasional + Hombres$Exfumador + Hombres$No_fumador + Hombres$Alcohol + Hombres$Can + Hombres$Co + Hombres$Hip + Hombres$Ejercicio_0 + Hombres$Ejercicio_1_2 + Hombres$Ejercicio_3_4 + Hombres$Ejercicio_5_6 + Hombres$Ejercicio_7 + Hombres$Insuficiente + Hombres$Normopeso + Hombres$Sobrepeso + Hombres$Obesidad + Hombres$Sueno + Hombres$Renta + Hombres$Edad, data= Hombres,family = Gamma(link = "inverse"))
```

```
#summary(modelo_0) #Este modelo debería ser el básico, el resto sale a partir de este
```

```
modelo_1=glm(Hombres$X2020 ~ Hombres$Fumador_diario:Hombres$Ca + Hombres$Ejercicio_1_2 + Hombres$Sobrepeso + Hombres$Sueno + Hombres$Edad, data= Hombres,family = Gamma(link = "inverse"))
```

```
summary(modelo_1)
```

```
modelo_2=glm(Hombres$X2020 ~ Hombres$Exfumador:Hombres$Hip + Hombres$Ejercicio_0 : Hombres$Insuficiente + Hombres$Sueno + Hombres$Edad, data= Hombres,family = Gamma(link = "inverse"))
```

```
summary(modelo_2)
```

```
modelo_3=glm(Hombres$X2020 ~ Hombres$Alcohol + Hombres$Edad, data= Hombres,family
= Gamma(link = "inverse"))
```

```
summary(modelo_3)
```

```
modelo_4=glm(Hombres$X2020 ~ Hombres$No_fumador +
Hombres$Ejercicio_3_4:Hombres$Obesidad + Hombres$Renta +Hombres$Edad, data=
Hombres,family = Gamma(link = "inverse"))
```

```
summary(modelo_4)
```

```
modelo_5=glm(Hombres$X2020 ~ Hombres$Fumador_ocasional + Hombres$Hip +
Hombres$Ejercicio_5_6 + Hombres$Normopeso + Hombres$Sueno + Hombres$Edad, data=
Hombres,family = Gamma(link = "inverse"))
```

```
summary(modelo_5)
```

```
#####Mismos modelos para mujeres#####
```

```
modelo_6=glm(Mujeres$X2020 ~ Mujeres$Exfumador +
Mujeres$Ejercicio_1_2:Mujeres$Obesidad+ Mujeres$Renta + Mujeres$Edad, data=
Mujeres,family = Gamma(link = "inverse"))
```

```
summary(modelo_6)
```

```
modelo_7=glm(Mujeres$X2020 ~ Mujeres$Fumador_ocasional : Mujeres$Alcohol +
Mujeres$Ejercicio_0 : Mujeres$Normopeso + Mujeres$Edad, data= Mujeres,family =
Gamma(link = "inverse"))
```

```
summary(modelo_7)
```

```
modelo_8=glm(Mujeres$X2020 ~ Mujeres$Fumador_ocasional + Mujeres$Alcohol+
Mujeres$Can + Mujeres$Ejercicio_1_2:Mujeres$Sobrepeso + Mujeres$Edad, data=
Mujeres,family = Gamma(link = "inverse"))
```

```
summary(modelo_8)
```

```
modelo_9=glm(Mujeres$X2020 ~ Mujeres$Exfumador +
Mujeres$Ejercicio_5_6:Mujeres$Insuficiente + Mujeres$Edad, data= Mujeres,family =
Gamma(link = "inverse"))
```

```
summary(modelo_9)
```

```
modelo_10=glm(Mujeres$X2020 ~ Mujeres$Ejercicio_3_4:Mujeres$Edad, data=
Mujeres,family = Gamma(link = "inverse"))
```

```
summary(modelo_10)
```

```
#####Predicciones con los modelos#####
```

```
#Creamos los dataframe de cada modelo con los correspondientes factores de riesgo
```

```
Datos_prediccion_1=data.frame(Hombres$Fumador_diario*Hombres$Ca ,  
Hombres$Ejercicio_1_2 , Hombres$Sobrepeso , Hombres$Sueno , Hombres$Edad)
```

```
Datos_prediccion_2=data.frame(Hombres$Exfumador*Hombres$Hip , Hombres$Ejercicio_0 *  
Hombres$Insuficiente , Hombres$Sueno , Hombres$Edad)
```

```
Datos_prediccion_3=data.frame(Hombres$Alcohol , Hombres$Edad)
```

```
Datos_prediccion_4=data.frame(Hombres$No_fumador ,  
Hombres$Ejercicio_3_4*Hombres$Obesidad , Hombres$Renta , Hombres$Edad)
```

```
Datos_prediccion_5=data.frame(Hombres$Fumador_ocasional , Hombres$Hip ,  
Hombres$Ejercicio_5_6 , Hombres$Normopeso , Hombres$Sueno , Hombres$Edad)
```

```
Datos_prediccion_6=data.frame(Mujeres$Exfumador ,  
Mujeres$Ejercicio_1_2*Mujeres$Obesidad , Mujeres$Renta , Mujeres$Edad)
```

```
Datos_prediccion_7=data.frame(Mujeres$Fumador_ocasional*Mujeres$Alcohol ,  
Mujeres$Ejercicio_0 * Mujeres$Normopeso , Mujeres$Edad)
```

```
Datos_prediccion_8=data.frame(Mujeres$Fumador_ocasional , Mujeres$Alcohol , Mujeres$Can  
 , Mujeres$Ejercicio_1_2*Mujeres$Sobrepeso , Mujeres$Edad)
```

```
Datos_prediccion_9=data.frame(Mujeres$Exfumador ,  
Mujeres$Ejercicio_5_6*Mujeres$Insuficiente, Mujeres$Edad)
```

```
Datos_prediccion_10=data.frame(Mujeres$Ejercicio_3_4*Mujeres$Edad)
```

```
#####Hacemos predicciones#####
```

```
prediccion_modelo_1 <- predict(modelo_1, newdata = Datos_prediccion_1, type = "response")
```

```
print(prediccion_modelo_1)
```

```
prediccion_modelo_1 = data.frame(prediccion_modelo_1)
```

```
prediccion_modelo_2 <- predict(modelo_2, newdata = Datos_prediccion_2, type = "response")
```

```
print(prediccion_modelo_2)
```

```
prediccion_modelo_2 = data.frame(prediccion_modelo_2)
```

```
prediccion_modelo_3 <- predict(modelo_3, newdata = Datos_prediccion_3, type = "response")
```

```
print(prediccion_modelo_3)
```

```
prediccion_modelo_3 = data.frame(prediccion_modelo_3)
```



```

prediccion_modelo_4 <- predict(modelo_4, newdata = Datos_prediccion_4, type = "response")
print(prediccion_modelo_4)
prediccion_modelo_4 = data.frame(prediccion_modelo_4)
prediccion_modelo_5 <- predict(modelo_5, newdata = Datos_prediccion_5, type = "response")
print(prediccion_modelo_5)
prediccion_modelo_5 = data.frame(prediccion_modelo_5)
prediccion_modelo_6 <- predict(modelo_6, newdata = Datos_prediccion_6, type = "response")
print(prediccion_modelo_6)
prediccion_modelo_6 = data.frame(prediccion_modelo_6)
prediccion_modelo_7 <- predict(modelo_7, newdata = Datos_prediccion_7, type = "response")
print(prediccion_modelo_7)
prediccion_modelo_7 = data.frame(prediccion_modelo_7)
prediccion_modelo_8 <- predict(modelo_8, newdata = Datos_prediccion_8, type = "response")
print(prediccion_modelo_8)
prediccion_modelo_8 = data.frame(prediccion_modelo_8)
prediccion_modelo_9 <- predict(modelo_9, newdata = Datos_prediccion_9, type = "response")
print(prediccion_modelo_9)
prediccion_modelo_9 = data.frame(prediccion_modelo_9)
prediccion_modelo_10 <- predict(modelo_10, newdata = Datos_prediccion_10, type =
"response")
print(prediccion_modelo_10)
prediccion_modelo_10 = data.frame(prediccion_modelo_10)
predicciones_totales=data.frame(as.vector(seq(from=11, to=64,
by=1)),prediccion_modelo_1,prediccion_modelo_2,prediccion_modelo_3,prediccion_modelo_4
,prediccion_modelo_5,prediccion_modelo_6,prediccion_modelo_7,prediccion_modelo_8,predic
cion_modelo_9,prediccion_modelo_10)
write.table(predicciones_totales,"predicciones_totales.txt", sep="\t", dec=",", col.names =
TRUE)

```

```
#####MODELO LEE-CARTER Y PROYECCIONES#####
```

```
Directorio2="C:/Users/ricar/OneDrive/Escritorio/Máster/TFM/Base de  
Datos/ESP/ESP/STATS/"
```

```
setwd(Directorio2)
```

```
Mortalidad_prueba <-  
read.demogdata("Mx_1x1_1960_2020_11_64.txt","Exposures_1x1_1960_2020_11_64.txt",  
type="mortality", label="Spain")
```

```
###MODELAMOS Y PREDECIMOS PARA MUJERES
```

```
modeloLC_mujeres=lca(data=Mortalidad_prueba,series=names(Mortalidad_prueba$rate)[1],  
years=Mortalidad_prueba$year,ages=Mortalidad_prueba$age, adjust = "dt")
```

```
summary(modeloLC_mujeres)
```

```
plot(modeloLC_mujeres)
```

```
plot(residuals(modeloLC_mujeres))
```

```
prediccion_modeloLC_mujeres=forecast(modeloLC_mujeres,h = 10, se = "innovdrift",  
jumpchoice ="actual",level = 95)
```

```
summary(prediccion_modeloLC_mujeres)
```

```
plot(prediccion_modeloLC_mujeres, plot.type = c("function", "component", "variance"))
```

```
# Extraer los valores de la tasa de mortalidad predicha
```

```
tasa_mortalidad_predicha_mujeres <- prediccion_modeloLC_mujeres$rate$female
```

```
plot(tasa_mortalidad_predicha_mujeres[,10])
```

```
# Imprimir los valores
```

```
print("Tasa de mortalidad predicha_mujeres:")
```

```
print(tasa_mortalidad_predicha_mujeres)
```

```
# Crear un data frame con los valores de la predicción
```

```
tabla_prediccion_mujeres <- data.frame(Tasa_predicha_mujeres =  
tasa_mortalidad_predicha_mujeres)
```

```
tabla_prediccion_mujeres <- tabla_prediccion_mujeres[,10]
```

```
# Imprimir la tabla
```

```

print(tabla_prediccion_mujeres)

tabla_prediccion_mujeres= data.frame(tabla_prediccion_mujeres)

####MODELAMOS Y PREDECIMOS PARA HOMBRES

modeloLC_hombres=lca(data=Mortalidad_prueba,series=names(Mortalidad_prueba$rate)[2],
years=Mortalidad_prueba$year,ages=Mortalidad_prueba$age, adjust = "dt")

summary(modeloLC_hombres)

plot(modeloLC_hombres)

plot(residuals(modeloLC_hombres))

prediccion_modeloLC_hombres=forecast(modeloLC_hombres,h = 10, se = "innovdrift",
jumpchoice ="actual",level = 95)

summary(prediccion_modeloLC_hombres)

plot(prediccion_modeloLC_hombres, plot.type = c("function", "component", "variance"))

# Extraer los valores de la tasa de mortalidad predicha

tasa_mortalidad_predicha_hombres <- prediccion_modeloLC_hombres$rate$male

plot(tasa_mortalidad_predicha_hombres[,10])

# Imprimir los valores

print("Tasa de mortalidad predicha_hombres:")

print(tasa_mortalidad_predicha_hombres)

# Crear un data frame con los valores de la predicción

tabla_prediccion_hombres <- data.frame(Tasa_predicha_hombres =
tasa_mortalidad_predicha_hombres)

tabla_prediccion_hombres <- tabla_prediccion_hombres[,10]

# Imprimir la tabla

print(tabla_prediccion_hombres)

tabla_prediccion_hombres=data.frame(tabla_prediccion_hombres)

#####Comprobamos si las predicciones son correctas#####

Directorio2="C:/Users/ricar/OneDrive/Escritorio/Máster/TFM/Base
Datos/ESP/ESP/STATS/"
de

```

```

setwd(Directorio2)

datos_backtesting <- read.demogdata("Mx_1x1_1960_2020_11_64.txt", "Exposures_1x1_1960_2020_11_64.txt",
type="mortality", label="Spain")

# Definimos los años de entrenamiento y prueba
train_years <- 1960:2010
test_years <- 2011:2020

# Creamos conjuntos de entrenamiento y prueba
train_data <- extract.years(datos_backtesting, train_years)
test_data <- extract.years(datos_backtesting, test_years)

lc_fit_train <- lca(train_data, series = "total")
summary(lc_fit_train)

# Proyección para el período de prueba
lc_forecast <- forecast(lc_fit_train, h = length(test_years))
summary(lc_forecast)
plot(lc_forecast)

proyeccion_lc_2020=lc_forecast$rate$total[,10] #tasa real proyectada a 2020
ts.plot(proyeccion_lc_2020)

real_2020=test_data$rate$total[,10]
ts.plot(real_2020)

# Generamos el eje X
x <- 11:64

# Dibujamos la serie de tiempo
plot(x, proyeccion_lc_2020, type = "o", col = "blue", xlab = "Edades", ylab = "Tasas de
mortalidad", main = "Nivel de ajuste backtesting",
ylim = range(c(proyeccion_lc_2020, real_2020)))
lines(x, real_2020, type = "o", col = "red")
legend("topright", legend = c("Tasa proyectada a 2020", "Tasa real de 2020"), col = c("blue",
"red"), lty = 1, pch = 1)

```

```

pre=test_data$rate$total[,20]

post=lc_forecast$rate$total[,20]

diferencia_prepost=post-pre

max(diferencia_prepost)

min(diferencia_prepost)

#####PROYECTAMOS LOS MODELOS OBTENIDOS HACIA 2030#####

#Modelo 1

datosparaproyeccion_modelo_1 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_1<-
datosparaproyeccion_modelo_1[datosparaproyeccion_modelo_1$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

año2020_modelo_1$Male=as.character(prediccion_modelo_1[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_1[datosparaproyeccion_modelo_1$Year==2020]=año2020_mode
lo_1 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_1=data.frame(datosparaproyeccion_modelo_1) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_1$columna = paste(datosparaproyeccion_modelo_1$Year,
datosparaproyeccion_modelo_1$Age, datosparaproyeccion_modelo_1$Female,
datosparaproyeccion_modelo_1$Male, datosparaproyeccion_modelo_1$total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_1$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_1.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

datos_proyeccion_modelo_1=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_1.txt","Ex
posures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

```

```

modeloLC_1=lca(data=datos_proyeccion_modelo_1,series=names(datos_proyeccion_modelo_1
$rate)[2], years=datos_proyeccion_modelo_1$year,ages=datos_proyeccion_modelo_1$age,
adjust = "dt")

summary(modeloLC_1) # Aplicamos el modelo Lee-Carter

plot(modeloLC_1)

plot(residuals(modeloLC_1))

prediccion_modeloLC_1=forecast(modeloLC_1,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_1)

plot(prediccion_modeloLC_1, plot.type = c("function", "component", "variance")) # Graficamos
tasa_mortalidad_predicha_modelo_1 <- prediccion_modeloLC_1$rate$male # Extraemos solo
las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_1:")

print(tasa_mortalidad_predicha_modelo_1)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_1 <- data.frame(Tasa_predicha_modelo_1 =
tasa_mortalidad_predicha_modelo_1)

tabla_prediccion_modelo_1 <- tabla_prediccion_modelo_1[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_1)

#tabla_prediccion_modelo_1= data.frame(tabla_prediccion_modelo_1) # quizá no sea necesario

#Modelo 2

datosparaproyeccion_modelo_2 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_2<-
datosparaproyeccion_modelo_2[datosparaproyeccion_modelo_2$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

```

```

año2020_modelo_2$Male=as.character(prediccion_modelo_2[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_2[datosparaproyeccion_modelo_2$Year==2020]=año2020_modelo_2 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_2=data.frame(datosparaproyeccion_modelo_2) # Lo convertimos en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_2$columna = paste(datosparaproyeccion_modelo_2$Year,
datosparaproyeccion_modelo_2$Age,
datosparaproyeccion_modelo_2$Female,
datosparaproyeccion_modelo_2$Male, datosparaproyeccion_modelo_2$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_2$columna # Sustituimos la mortalidad que había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_2.txt", sep = "\t") # Lo guardamos como .txt para introducirlo después como demogdata

datos_proyeccion_modelo_2=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_2.txt","Exposures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto demogdata con la mortalidad calculada y la población del HMD

modeloLC_2=lca(data=datos_proyeccion_modelo_2,series=names(datos_proyeccion_modelo_2)$rate)[2], years=datos_proyeccion_modelo_2$year,ages=datos_proyeccion_modelo_2$age, adjust = "dt")

summary(modeloLC_2) # Aplicamos el modelo Lee-Carter

plot(modeloLC_2)

prediccion_modeloLC_2=forecast(modeloLC_2,h = 10, se = "innovdrift", jumpchoice ="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_2)

plot(prediccion_modeloLC_2, plot.type = c("function", "component", "variance")) # Graficamos tasa_mortalidad_predicha_modelo_2 <- prediccion_modeloLC_2$rate$male # Extraemos solo las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_2:")

```

```

print(tasa_mortalidad_predicha_modelo_2)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_2 <- data.frame(Tasa_predicha_modelo_2 =
tasa_mortalidad_predicha_modelo_2)

tabla_prediccion_modelo_2 <- tabla_prediccion_modelo_2[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_2)

#tabla_prediccion_modelo_2= data.frame(tabla_prediccion_modelo_2) # quizá no sea necesario

#Modelo 3

datosparaproyeccion_modelo_3 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_3<-
datosparaproyeccion_modelo_3[datosparaproyeccion_modelo_3$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

año2020_modelo_3$Male=as.character(prediccion_modelo_3[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_3[datosparaproyeccion_modelo_3$Year==2020]=año2020_mode
lo_3 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_3=data.frame(datosparaproyeccion_modelo_3) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_3$columna = paste(datosparaproyeccion_modelo_3$Year,
datosparaproyeccion_modelo_3$Age, datosparaproyeccion_modelo_3$Female,
datosparaproyeccion_modelo_3$Male, datosparaproyeccion_modelo_3$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_3$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_3.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

```



```

datos_proyeccion_modelo_3=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_3.txt","Ex
posures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

modeloLC_3=lca(data=datos_proyeccion_modelo_3,series=names(datos_proyeccion_modelo_3
$rate)[2], years=datos_proyeccion_modelo_3$year,ages=datos_proyeccion_modelo_3$age,
adjust = "dt")

summary(modeloLC_3) # Aplicamos el modelo Lee-Carter

plot(modeloLC_3)

prediccion_modeloLC_3=forecast(modeloLC_3,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_3)

plot(prediccion_modeloLC_3, plot.type = c("function", "component", "variance")) # Graficamos
tasa_mortalidad_predicha_modelo_3 <- prediccion_modeloLC_3$rate$male # Extraemos solo
las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_3:")

print(tasa_mortalidad_predicha_modelo_3)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_3 <- data.frame(Tasa_predicha_modelo_3 =
tasa_mortalidad_predicha_modelo_3)

tabla_prediccion_modelo_3 <- tabla_prediccion_modelo_3[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_3)

#tabla_prediccion_modelo_3= data.frame(tabla_prediccion_modelo_3) # quizá no sea necesario

#Modelo 4

datosparaproyeccion_modelo_4 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_4<-
datosparaproyeccion_modelo_4[datosparaproyeccion_modelo_4$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

```

```

año2020_modelo_4$Male=as.character(prediccion_modelo_4[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_4[datosparaproyeccion_modelo_4$Year==2020]=año2020_modelo_4 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_4=data.frame(datosparaproyeccion_modelo_4) # Lo convertimos en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_4$columna = paste(datosparaproyeccion_modelo_4$Year,
datosparaproyeccion_modelo_4$Age,
datosparaproyeccion_modelo_4$Female,
datosparaproyeccion_modelo_4$Male, datosparaproyeccion_modelo_4$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_4$columna # Sustituimos la mortalidad que había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_4.txt", sep = "\t") # Lo guardamos como .txt para introducirlo después como demogdata

datos_proyeccion_modelo_4=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_4.txt","Exposures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto demogdata con la mortalidad calculada y la población del HMD

modeloLC_4=lca(data=datos_proyeccion_modelo_4,series=names(datos_proyeccion_modelo_4)$rate)[2], years=datos_proyeccion_modelo_4$year,ages=datos_proyeccion_modelo_4$age, adjust = "dt")

summary(modeloLC_4) # Aplicamos el modelo Lee-Carter

plot(modeloLC_4)

prediccion_modeloLC_4=forecast(modeloLC_4,h = 10, se = "innovdrift", jumpchoice ="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_4)

plot(prediccion_modeloLC_4, plot.type = c("function", "component", "variance")) # Graficamos tasa_mortalidad_predicha_modelo_4 <- prediccion_modeloLC_4$rate$male # Extraemos solo las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_4:")

```

```

print(tasa_mortalidad_predicha_modelo_4)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_4 <- data.frame(Tasa_predicha_modelo_4 =
tasa_mortalidad_predicha_modelo_4)

tabla_prediccion_modelo_4 <- tabla_prediccion_modelo_4[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_4)

#tabla_prediccion_modelo_4= data.frame(tabla_prediccion_modelo_4) # quizá no sea necesario

#Modelo 5

datosparaproyeccion_modelo_5 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_5<-
datosparaproyeccion_modelo_5[datosparaproyeccion_modelo_5$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

año2020_modelo_5$Male=as.character(prediccion_modelo_5[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_5[datosparaproyeccion_modelo_5$Year==2020]=año2020_mode
lo_5 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_5=data.frame(datosparaproyeccion_modelo_5) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_5$columna = paste(datosparaproyeccion_modelo_5$Year,
datosparaproyeccion_modelo_5$Age, datosparaproyeccion_modelo_5$Female,
datosparaproyeccion_modelo_5$Male, datosparaproyeccion_modelo_5$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_5$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_5.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

```

```

datos_proyeccion_modelo_5=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_5.txt","Ex
posures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

modeloLC_5=lca(data=datos_proyeccion_modelo_5,series=names(datos_proyeccion_modelo_5
$rate)[2], years=datos_proyeccion_modelo_5$year,ages=datos_proyeccion_modelo_5$age,
adjust = "dt")

summary(modeloLC_5) # Aplicamos el modelo Lee-Carter

plot(modeloLC_5)

prediccion_modeloLC_5=forecast(modeloLC_5,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_5)

plot(prediccion_modeloLC_5, plot.type = c("function", "component", "variance")) # Graficamos
tasa_mortalidad_predicha_modelo_5 <- prediccion_modeloLC_5$rate$male # Extraemos solo
las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_5:")

print(tasa_mortalidad_predicha_modelo_5)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_5 <- data.frame(Tasa_predicha_modelo_5 =
tasa_mortalidad_predicha_modelo_5)

tabla_prediccion_modelo_5 <- tabla_prediccion_modelo_5[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_5)

#tabla_prediccion_modelo_5= data.frame(tabla_prediccion_modelo_5) # quizá no sea necesario

#Modelo 6

datosparaproyeccion_modelo_6 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_6<-
datosparaproyeccion_modelo_6[datosparaproyeccion_modelo_6$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

```

```

año2020_modelo_6$Female=as.character(prediccion_modelo_6[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_6[datosparaproyeccion_modelo_6$Year==2020]=año2020_modelo_6 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_6=data.frame(datosparaproyeccion_modelo_6) # Lo convertimos en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_6$columna = paste(datosparaproyeccion_modelo_6$Year,
datosparaproyeccion_modelo_6$Age,
datosparaproyeccion_modelo_6$Female,
datosparaproyeccion_modelo_6$Male, datosparaproyeccion_modelo_6$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_6$columna # Sustituimos la mortalidad que había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_6.txt", sep = "\t") # Lo guardamos como .txt para introducirlo después como demogdata

datos_proyeccion_modelo_6=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_6.txt","Exposures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto demogdata con la mortalidad calculada y la población del HMD

modeloLC_6=lca(data=datos_proyeccion_modelo_6,series=names(datos_proyeccion_modelo_6)$rate)[1], years=datos_proyeccion_modelo_6$year,ages=datos_proyeccion_modelo_6$age, adjust = "dt")

summary(modeloLC_6) # Aplicamos el modelo Lee-Carter

plot(modeloLC_6)

prediccion_modeloLC_6=forecast(modeloLC_6,h = 10, se = "innovdrift", jumpchoice ="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_6)

plot(prediccion_modeloLC_6, plot.type = c("function", "component", "variance")) # Graficamos tasa_mortalidad_predicha_modelo_6 <- prediccion_modeloLC_6$rate$female # Extraemos solo las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_6:")

```

```

print(tasa_mortalidad_predicha_modelo_6)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_6 <- data.frame(Tasa_predicha_modelo_6 =
tasa_mortalidad_predicha_modelo_6)

tabla_prediccion_modelo_6 <- tabla_prediccion_modelo_6[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_6)

#tabla_prediccion_modelo_6= data.frame(tabla_prediccion_modelo_6) # quizá no sea necesario

#Modelo 7

datosparaproyeccion_modelo_7 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_7<-
datosparaproyeccion_modelo_7[datosparaproyeccion_modelo_7$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

año2020_modelo_7$Female=as.character(prediccion_modelo_7[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_7[datosparaproyeccion_modelo_7$Year==2020]=año2020_mode
lo_7 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_7=data.frame(datosparaproyeccion_modelo_7) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_7$columna = paste(datosparaproyeccion_modelo_7$Year,
datosparaproyeccion_modelo_7$Age, datosparaproyeccion_modelo_7$Female,
datosparaproyeccion_modelo_7$Male, datosparaproyeccion_modelo_7$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_7$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_7.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

```

```

datos_proyeccion_modelo_7=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_7.txt","Ex
posures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

modeloLC_7=lca(data=datos_proyeccion_modelo_7,series=names(datos_proyeccion_modelo_7
$rate)[1], years=datos_proyeccion_modelo_7$year,ages=datos_proyeccion_modelo_7$age,
adjust = "dt")

summary(modeloLC_7) # Aplicamos el modelo Lee-Carter

plot(modeloLC_7)

prediccion_modeloLC_7=forecast(modeloLC_7,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_7)

plot(prediccion_modeloLC_7, plot.type = c("function", "component", "variance")) # Graficamos
tasa_mortalidad_predicha_modelo_7 <- prediccion_modeloLC_7$rate$female # Extraemos solo
las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_7:")

print(tasa_mortalidad_predicha_modelo_7)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_7 <- data.frame(Tasa_predicha_modelo_7 =
tasa_mortalidad_predicha_modelo_7)

tabla_prediccion_modelo_7 <- tabla_prediccion_modelo_7[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_7)

#tabla_prediccion_modelo_7= data.frame(tabla_prediccion_modelo_7) # quizá no sea necesario

#Modelo 8

datosparaproyeccion_modelo_8 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_8<-
datosparaproyeccion_modelo_8[datosparaproyeccion_modelo_8$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

```

```

año2020_modelo_8$Female=as.character(prediccion_modelo_8[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_8[datosparaproyeccion_modelo_8$Year==2020]=año2020_modelo_8
# Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_8=data.frame(datosparaproyeccion_modelo_8) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_8$columna = paste(datosparaproyeccion_modelo_8$Year,
datosparaproyeccion_modelo_8$Age,
datosparaproyeccion_modelo_8$Female,
datosparaproyeccion_modelo_8$Male, datosparaproyeccion_modelo_8$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_8$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_8.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

datos_proyeccion_modelo_8=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_8.txt","Ex
posures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

modeloLC_8=lca(data=datos_proyeccion_modelo_8,series=names(datos_proyeccion_modelo_8
$rate)[1], years=datos_proyeccion_modelo_8$year,ages=datos_proyeccion_modelo_8$age,
adjust = "dt")

summary(modeloLC_8) # Aplicamos el modelo Lee-Carter

plot(modeloLC_8)

prediccion_modeloLC_8=forecast(modeloLC_8,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_8)

plot(prediccion_modeloLC_8, plot.type = c("function", "component", "variance")) # Graficamos
tasa_mortalidad_predicha_modelo_8 <- prediccion_modeloLC_8$rate$female # Extraemos solo
las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_8:")

```



```

print(tasa_mortalidad_predicha_modelo_8)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_8 <- data.frame(Tasa_predicha_modelo_8 =
tasa_mortalidad_predicha_modelo_8)

tabla_prediccion_modelo_8 <- tabla_prediccion_modelo_8[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_8)

#tabla_prediccion_modelo_8= data.frame(tabla_prediccion_modelo_8) # quizá no sea necesario

#Modelo 9

datosparaproyeccion_modelo_9 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_9<-
datosparaproyeccion_modelo_9[datosparaproyeccion_modelo_9$Year==2020] # Extraemos los
datos que pertenezcan únicamente a 2020

año2020_modelo_9$Female=as.character(prediccion_modelo_9[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_9[datosparaproyeccion_modelo_9$Year==2020]=año2020_mode
lo_9 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_9=data.frame(datosparaproyeccion_modelo_9) # Lo convertimos
en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_9$columna = paste(datosparaproyeccion_modelo_9$Year,
datosparaproyeccion_modelo_9$Age, datosparaproyeccion_modelo_9$Female,
datosparaproyeccion_modelo_9$Male, datosparaproyeccion_modelo_9$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_9$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_9.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

```

```
datos_proyeccion_modelo_9=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_9.txt","Exposures_1x1_1960_2020_11_64.txt", type="mortality", label ="Spain") # Creamos el objeto demogdata con la mortalidad calculada y la población del HMD
```

```
modeloLC_9=lca(data=datos_proyeccion_modelo_9,series=names(datos_proyeccion_modelo_9$rate)[1], years=datos_proyeccion_modelo_9$year,ages=datos_proyeccion_modelo_9$age, adjust = "dt")
```

```
summary(modeloLC_9) # Aplicamos el modelo Lee-Carter
```

```
plot(modeloLC_9)
```

```
prediccion_modeloLC_9=forecast(modeloLC_9,h = 10, se = "innovdrift", jumpchoice ="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030
```

```
summary(prediccion_modeloLC_9)
```

```
plot(prediccion_modeloLC_9, plot.type = c("function", "component", "variance")) # Graficamos
```

```
tasa_mortalidad_predicha_modelo_9 <- prediccion_modeloLC_9$rate$female # Extraemos solo las tasas para los hombres
```

```
# Imprimimos los valores
```

```
print("tasa_mortalidad_predicha_modelo_9:")
```

```
print(tasa_mortalidad_predicha_modelo_9)
```

```
# Creamos un df con los valores de la predicción
```

```
tabla_prediccion_modelo_9 <- data.frame(Tasa_predicha_modelo_9 = tasa_mortalidad_predicha_modelo_9)
```

```
tabla_prediccion_modelo_9 <- tabla_prediccion_modelo_9[,10] # Extraemos solo las tasas del año 2030
```

```
# Imprimimos la tabla
```

```
print(tabla_prediccion_modelo_9)
```

```
#tabla_prediccion_modelo_9= data.frame(tabla_prediccion_modelo_9) # quizá no sea necesario
```

```
#Modelo 10
```

```

datosparaproyeccion_modelo_10 <- fread("Mx_1x1_1960_2020_11_64.txt", sep = " ", header =
TRUE) # Importamos la base de datos original (creada una copia por si me confundo)

año2020_modelo_10<-
datosparaproyeccion_modelo_10[datosparaproyeccion_modelo_10$Year==2020] # Extraemos
los datos que pertenezcan únicamente a 2020

año2020_modelo_10$Female=as.character(prediccion_modelo_10[,1]) # De los datos de 2020,
extraemos solo los de los hombres

datosparaproyeccion_modelo_10[datosparaproyeccion_modelo_10$Year==2020]=año2020_mo
delo_10 # Y los sustituimos por los que hemos predicho en el GLM

datosparaproyeccion_modelo_10=data.frame(datosparaproyeccion_modelo_10) # Lo
convertimos en df y lo juntamos en una sola columna, para adaptar el df al tipo demogdata

datosparaproyeccion_modelo_10$columna = paste(datosparaproyeccion_modelo_10$Year,
datosparaproyeccion_modelo_10$Age, datosparaproyeccion_modelo_10$Female,
datosparaproyeccion_modelo_10$Male, datosparaproyeccion_modelo_10$Total., sep = " ")

plantilla=fread("Mx_1x1_1960_2020_11_64.txt", sep = "\t", header = TRUE) # Importamos de
nuevo la base de datos (por si hubiera algún problema)

plantilla[3:3296,1]=datosparaproyeccion_modelo_10$columna # Sustituimos la mortalidad que
había originalmente por el df que acabamos de crear

fwrite(plantilla, "Mx_1x1_1960_2020_11_64_modelo_10.txt", sep = "\t") # Lo guardamos como
.txt para introducirlo después como demogdata

datos_proyeccion_modelo_10=read.demogdata("Mx_1x1_1960_2020_11_64_modelo_10.txt",
Exposures_1x1_1960_2020_11_64.txt", type="mortality", label="Spain") # Creamos el objeto
demogdata con la mortalidad calculada y la población del HMD

modeloLC_10=lca(data=datos_proyeccion_modelo_10,series=names(datos_proyeccion_modelo
_10$rate)[1],
years=datos_proyeccion_modelo_10$year,ages=datos_proyeccion_modelo_10$age, adjust =
"dt")

summary(modeloLC_10) # Aplicamos el modelo Lee-Carter

plot(modeloLC_10)

prediccion_modeloLC_10=forecast(modeloLC_10,h = 10, se = "innovdrift", jumpchoice
="actual",level = 95) # Hacemos una proyección de la mortalidad calculada hasta 2030

summary(prediccion_modeloLC_10)

```

```

plot(prediccion_modeloLC_10, plot.type = c("function", "component", "variance")) #
Graficamos

tasa_mortalidad_predicha_modelo_10 <- prediccion_modeloLC_10$rate$female # Extraemos
solo las tasas para los hombres

# Imprimimos los valores

print("tasa_mortalidad_predicha_modelo_10:")

print(tasa_mortalidad_predicha_modelo_10)

# Creamos un df con los valores de la predicción

tabla_prediccion_modelo_10 <- data.frame(Tasa_predicha_modelo_10 =
tasa_mortalidad_predicha_modelo_10)

tabla_prediccion_modelo_10 <- tabla_prediccion_modelo_10[,10] # Extraemos solo las tasas del
año 2030

# Imprimimos la tabla

print(tabla_prediccion_modelo_10)

#tabla_prediccion_modelo_10= data.frame(tabla_prediccion_modelo_10) # quizá no sea
necesario

####Ordenamos las proyecciones, separando entre hombres y mujeres

edades=11:64

# Graficamos las 6 series de tiempo de los hombres

proyecciones_hombres=data.frame(edades,tabla_prediccion_hombres,tabla_prediccion_modelo
_1,tabla_prediccion_modelo_2,tabla_prediccion_modelo_3,tabla_prediccion_modelo_4,tabla_p
rediccion_modelo_5)

write.table(proyecciones_hombres, "proyecciones_hombres.txt", sep="\t", dec = ",", col.names =
TRUE)

ggplot(data = proyecciones_hombres, aes(x = edades)) +
  geom_line(aes(y = tabla_prediccion_hombres, color = "Proyección original")) +
  geom_line(aes(y = tabla_prediccion_modelo_1, color = "Modelo 1")) +
  geom_line(aes(y = tabla_prediccion_modelo_2, color = "Modelo 2")) +
  geom_line(aes(y = tabla_prediccion_modelo_3, color = "Modelo 3")) +
  geom_line(aes(y = tabla_prediccion_modelo_4, color = "Modelo 4")) +

```

```

geom_line(aes(y = tabla_prediccion_modelo_5, color = "Modelo 5")) +
labs(color = "Modelos proyectados a 2030") +
theme_minimal()

# Graficamos las 6 series de tiempo de las mujeres
proyecciones_mujeres=data.frame(edades,tabla_prediccion_mujeres,tabla_prediccion_modelo_
6,tabla_prediccion_modelo_7,tabla_prediccion_modelo_8,tabla_prediccion_modelo_9,tabla_pr
ediccion_modelo_10)

write.table(proyecciones_mujeres, "proyecciones_mujeres.txt", sep="\t", dec = ",", col.names =
TRUE)

ggplot(data = proyecciones_mujeres, aes(x = edades)) +
  geom_line(aes(y = tabla_prediccion_mujeres, color = "Proyección original")) +
  geom_line(aes(y = tabla_prediccion_modelo_6, color = "Modelo 6")) +
  geom_line(aes(y = tabla_prediccion_modelo_7, color = "Modelo 7")) +
  geom_line(aes(y = tabla_prediccion_modelo_8, color = "Modelo 8")) +
  geom_line(aes(y = tabla_prediccion_modelo_9, color = "Modelo 9")) +
  geom_line(aes(y = tabla_prediccion_modelo_10, color = "Modelo 10")) +
  labs(color = "Modelos proyectados a 2030") +
  theme_minimal()

# Creamos un df con todas las mortalidades predichas con la original proyectada para hombres
comparación_hombres=data.frame(edades,
                                tabla_prediccion_hombres,
                                tabla_prediccion_modelo_1,tabla_prediccion_modelo_2,tabla_pred
iccion_modelo_3,tabla_pred
iccion_modelo_4,tabla_prediccion_modelo_5)

print(comparación_hombres)

comparación_hombres_difs_1=tabla_prediccion_modelo_1-tabla_prediccion_hombres
comparación_hombres_difs_2=tabla_prediccion_modelo_2-tabla_prediccion_hombres
comparación_hombres_difs_3=tabla_prediccion_modelo_3-tabla_prediccion_hombres
comparación_hombres_difs_4=tabla_prediccion_modelo_4-tabla_prediccion_hombres
comparación_hombres_difs_5=tabla_prediccion_modelo_5-tabla_prediccion_hombres

# Creamos un df con todas las mortalidades predichas con la original proyectada para mujeres

```

```

comparación_mujeres=data.frame(edades,
                                tabla_prediccion_mujeres,
                                tabla_prediccion_modelo_6,tabla_prediccion_modelo_7,tabla_prediccion_modelo_8,tabla_prediccion_modelo_9,tabla_prediccion_modelo_10)

print(comparación_mujeres)

comparación_mujeres_difs_1=tabla_prediccion_modelo_6-tabla_prediccion_mujeres
comparación_mujeres_difs_2=tabla_prediccion_modelo_7-tabla_prediccion_mujeres
comparación_mujeres_difs_3=tabla_prediccion_modelo_8-tabla_prediccion_mujeres
comparación_mujeres_difs_4=tabla_prediccion_modelo_9-tabla_prediccion_mujeres
comparación_mujeres_difs_5=tabla_prediccion_modelo_10-tabla_prediccion_mujeres

comparación_hombres_difs=data.frame(edades,comparación_hombres_difs_1,comparación_hombres_difs_2,comparación_hombres_difs_3,comparación_hombres_difs_4,comparación_hombres_difs_5)

print(comparación_hombres_difs)

comparación_mujeres_difs=data.frame(edades,comparación_mujeres_difs_1,comparación_mujeres_difs_2,comparación_mujeres_difs_3,comparación_mujeres_difs_4,comparación_mujeres_difs_5)

print(comparación_mujeres_difs)

write.table(comparación_hombres_difs, "comparación_hombres_difs.txt", sep="\t", dec = ",", col.names = TRUE)

write.table(comparación_mujeres_difs, "comparación_mujeres_difs.txt", sep="\t", dec = ",", col.names = TRUE)

```