

GRADO: ECONOMÍA

Curso 2017/2018

ANÁLISIS DE DETECCIÓN TEMPRANA DEL CÁNCER DE MAMA

Autora: Olatz Ordiales Cuesta

Directoras: Ana Fernández Sainz

Ana María Martín Arroyuelos

Bilbao, a 28 de junio de 2018



ÍNDICE

Resumen	5
Abstract.....	5
Objetivos del trabajo	6
Estructura del trabajo.....	6
¿Por qué elegí este tema?	8
1. INTRODUCCIÓN	8
2. EL CÁNCER DE MAMA/SENO	10
2.1. EL CÁNCER DE MAMA A NIVEL MUNDIAL.....	12
2.2. EL CÁNCER DE MAMA EN ESTADOS UNIDOS.....	13
2.3. EL CÁNCER DE MAMA EN ESPAÑA.....	15
2.4. EL CÁNCER DE MAMA EN EL PAÍS VASCO.....	18
3. FACTORES SOCIODEMOGRÁFICOS QUE PUEDEN AFECTAR A LA ENFERMEDAD.....	19
4. ANÁLISIS ESTADÍSTICO Y ECONÓMICO DE CASOS DE CÁNCER DE MAMA.....	22
4.1. DESCRIPCIÓN DE LAS VARIABLES	23
4.2. CONTRASTES DE INDEPENDENCIA	25
• RAZA.....	27
• EDAD	29
• ESTADO CIVIL	29
• LATERALIDAD	30
• INTERPRETACIÓN DE RESULTADOS.....	31
4.3. ANÁLISIS DE REGRESIÓN	31
• ESTIMACIÓN POR MCO	34
○ Contraste de significación individual	35
○ Contraste de significación conjunta.....	37
• CONTRASTES DE HETEROCEDASTICIDAD	38
○ Contraste de White.....	39
○ Contraste de Breusch Y Pagan.....	39
• ESTIMACIÓN ROBUSTA POR MCO CON DESVIACIONES TÍPICAS ROBUSTAS	40
• INTERPRETACIÓN DE LOS PARÁMETROS	42
5. CONCLUSIONES.....	43
6. BIBLIOGRAFÍA	45

ÍNDICE DE FIGURAS, GRÁFICOS Y TABLAS

FIGURAS

Figura 1: Tasas de mortalidad del cáncer de mama a nivel mundial.....	13
Figura 2: Tasas de incidencia del cáncer de mama en Estados Unidos por estado.....	14
Figura 3: Tasa de incidencia del cáncer de mama en mujeres en Europa.....	15
Figura 4: Casos anuales de tumores malignos y defunciones por tipo de tumor y sexo.....	18
Figura 5: Diagrama de caja del tamaño del tumor.....	32
Figura 6: Diagrama del tamaño del tumor (Zoom de figura 5).....	32
Figura 7: Histograma del tamaño del tumor.....	35

GRÁFICOS

Gráfico 1: Distribución de la mortalidad por tipo de tumor en Estados Unidos.....	14
Gráfico 2: Distribución por tipo de tumor del cáncer en España en mujeres.....	16
Gráfico 3: Distribución de la mortalidad por cáncer en función del tipo de tumor.....	16
Gráfico 4: Distribución por edad del cáncer de mama en mujeres.....	17
Gráfico 5: Distribución por edad de la mortalidad de cáncer de mama.....	17
Gráfico 6: Distribución por raza de las mujeres con cáncer de mama.....	23
Gráfico 7: Distribución por estado civil de las mujeres con cáncer de mama.....	24
Gráfico 8: Distribución por edad de las mujeres con cáncer de mama.....	24
Gráfico 9: Distribución por grado de las mujeres con cáncer de mama.....	25

TABLAS

Tabla 1: Relación raza y grado del tumor.....	27
Tabla 2: Relación origen hispano y grado del tumor.....	28
Tabla 3: Relación edad y grado del tumor.....	29
Tabla 4: Relación estado civil y grado del tumor.....	29
Tabla 5: Relación lateralidad y grado del tumor.....	30
Tabla 6: Variables del modelo.....	33-34
Tabla 7: Estimación por MCO.....	36
Tabla 8: Contrastes de heterocedasticidad.....	39
Tabla 9: Estimación por MCO. Matriz de varianzas y covarianzas robustas.....	41

Resumen

El presente Trabajo de Fin de Grado trata, en primera instancia, de proporcionar una visión detallada sobre el cáncer de mama en mujeres, tanto a nivel mundial, como en Estados Unidos, España y País Vasco.

Además, se consideran las distintas variables sociodemográficas (raza, estado civil, origen hispano y edad) que pueden afectar a la tasa de incidencia del cáncer de mama, y se determina si hay alguna relación entre estas variables y el grado del tumor en el momento del diagnóstico. También se analiza a través de la especificación y estimación de un modelo econométrico, el efecto de dichas variables sobre el tamaño del tumor de las mujeres.

Los datos utilizados en el presente trabajo han sido obtenidos de la base de datos: SEER CANCER RESEARCH DATABASE, del Instituto Nacional de Cáncer de Estados Unidos (2017).

Palabras clave: cáncer, mujeres, contraste de independencia, estimaciones econométricas, MCO.

Abstract

This current Final Project deals firstly with a detailed view about breast cancer in women, explaining in detail the situation worldwide, as well as in USA, Spain and in the Basque Country.

In addition, the different sociodemographic variables (race, marital status, Hispanic origin and age) that can cause or affect the incidence rate of breast cancer are considered, and it is determined if there is any relationship between these variables and the grade of the tumour. Moreover, it is also analyzed the effect of those variables on the tumor size of women through the specification and estimation of an econometric model.

The data that has been used for this work have been obtained from a database from SEER CANCER RESEARCH DATABASE of the National Cancer Institute of the United States (2017).

Keywords: cancer, women, independence contrast, econometric estimation, OLS.

Objetivos del trabajo

Según la World Cancer Research Fund International (2012), el cáncer es una de las cuatro enfermedades, junto con las enfermedades cardiovasculares, las respiratorias y la diabetes, que constituyen el 80% de las enfermedades mortales en todo el mundo. Actualmente, en nuestra sociedad, una de cada ocho mujeres tendrá cáncer de mama a lo largo de su vida y, hoy en día, son muchas las mujeres que luchan por sobrevivir a un cáncer de mama. Según la Asociación Española Contra el Cáncer (2017), el cáncer de mama es el tumor más frecuente en las mujeres occidentales, por lo que resulta de gran importancia saber lo máximo posible sobre esta enfermedad.

En base a ello, los principales objetivos de este trabajo son:

- Entender el impacto que tiene esta enfermedad en la sociedad.
- Obtener una amplia visión sobre un tema que está siempre de actualidad y que moviliza ingentes recursos económicos y humanos.
- Analizar si existe relación entre determinadas variables sociodemográficas tales como el estado civil, la raza, el origen hispano, y la edad, con el grado y el tamaño del tumor en el momento del diagnóstico. El análisis se realiza para mujeres de Estados Unidos en el año 2014.

Estructura del trabajo

El estudio que se va a llevar a cabo consta de cuatro partes que se basan principalmente en un estudio teórico, y en un análisis estadístico y econométrico.

1. En la primera parte del trabajo se llevará a cabo un estudio teórico sobre el cáncer en general, fundamentalmente a partir de la información recogida en estudios de asociaciones contra el cáncer o diversas investigaciones que se han realizado sobre el mismo.

Se realizará una explicación más detallada acerca del cáncer de mama, explicando y analizando los diferentes síntomas, diagnósticos, tratamientos y factores a tener en cuenta para este estudio. Se analizarán los datos sobre el cáncer de mama en Estados Unidos, España y el País Vasco.

Por último, se hace referencia a la mejora del diagnóstico y cura del cáncer en el futuro, analizando cuáles son los avances que hay hoy en día en el cáncer de mama y resaltando

cuáles son las medidas llevadas a cabo por los expertos para poder decir que existen muchas expectativas acerca de la cura de esta enfermedad.

2. En la segunda parte del proyecto se revisan diferentes estudios ya realizados en torno al efecto que tienen sobre el diagnóstico del cáncer de mama las variables sociodemográficas consideradas.

Para llevar a cabo este análisis he utilizado principalmente artículos académicos y páginas web de instituciones que trabajan en el análisis y diagnóstico del cáncer.

3. En la tercera parte se realiza un análisis estadístico y econométrico del cáncer de mama en mujeres de Estados Unidos utilizando los datos que nos aporta SEER CANCER RESEARCH DATABASE (2017) del Instituto Nacional de Cáncer de Estados Unidos.

Para ello y, en primer lugar, se realiza un análisis descriptivo de las diferentes variables consideradas. Posteriormente, se realizan contrastes de independencia para saber qué variables son relevantes para la determinación del grado del cáncer de mama. Finalmente, se especifican y estiman distintos modelos de regresión lineales.

4. En la última parte del trabajo se exponen los principales resultados obtenidos gracias al estudio que se realiza en los apartados anteriores, lo que nos permitirá extraer las conclusiones del trabajo.

Como ya se ha dicho anteriormente, las fuentes de información para llevar a cabo el estudio han sido diversas. Elementos clave para la realización del proyecto han sido los datos obtenidos de SEER CANCER RESEARCH DATABASE (2017) del Instituto Nacional de Cáncer de Estados Unidos y la información recogida a través de un conjunto de varias asociaciones e institutos nacionales citados en la bibliografía.

Para realizar el análisis descriptivo y los contrastes de independencia he utilizado los programas SPSS (Versión 23.0.0.0) y Excel, y para estimar los distintos modelos de regresión, el programa Gretl (Versión 1.9.2).

Además, se han utilizado datos e información a la que se ha accedido a través de la búsqueda y lectura de noticias, artículos e informes de distintas revistas académicas que tratan del cáncer de mama, páginas de universidades o boletines del cáncer de mama o el cáncer en general.

¿Por qué elegí este tema?

Por último, el motivo que me ha llevado a realizar este Trabajo de Fin de Grado ha sido, principalmente, el gran interés personal por analizar el cáncer de mama en mujeres. Hoy en día sigue siendo una enfermedad para la que no se ha encontrado una cura definitiva y por la que están o estarán afectadas gran cantidad de mujeres que luchan diariamente por su vida después del diagnóstico. Todas conocemos en nuestro entorno mujeres diagnosticadas, algunas de las cuales han superado la enfermedad y otras, en cambio, se han quedado en el camino.

Además, este trabajo, me ha permitido aplicar los conocimientos adquiridos durante los estudios del Grado en Economía en las asignaturas de Estadística y Análisis de Datos, Estadística Aplicada a la Economía, Introducción a la Econometría y Econometría Aplicada, así como manejar herramientas útiles a la hora de procesar, analizar datos y obtener resultados.

Aunar ambos intereses me pareció la mejor forma de culminar mis estudios.

1. INTRODUCCIÓN

El cáncer es aún una enfermedad que a nivel mundial preocupa sobre manera, ya que todavía no se ha encontrado ninguna cura definitiva, aunque hay múltiples investigaciones abiertas desde diferentes perspectivas. El cáncer es en España la primera causa de muerte en hombres y la segunda en mujeres. Uno de cada tres ciudadanos españoles desarrollará algún tipo de cáncer antes de los 85 años (Borrás, 2015).

Aun siendo esta enfermedad muy importante, no es la única que necesita que se dedique una parte importante del PIB (Producto Interno Bruto) al gasto sanitario. Actualmente en España se destina a dicho concepto el 9,4% del PIB (Cabasés, 2015). Como todos conocemos, este porcentaje era más alto antes de la actual crisis económica, debido a ella se han hecho muchos recortes en sanidad, lo que ha provocado una reducción de las revisiones a los pacientes, un encarecimiento de muchos medicamentos y, además, una disminución en los gastos de investigación, acarreado en definitiva una sanidad de peor calidad. De hecho, los medicamentos oncológicos tienen precios muy elevados, lo que es resultado de la existencia de un mercado poco competitivo. Aun así, en España, se destinan 4.820 millones de euros al tratamiento del cáncer anualmente, lo que supone el 4,8% del PIB.

Es una de las enfermedades con mayor índice de mortalidad en el mundo, y la segunda causa de muerte. Según la Organización Mundial de la Salud (OMS), en el año 2015 murieron 8,8 millones

de personas por causa de esta enfermedad, lo que supone casi una de cada seis defunciones en el mundo.

El cáncer es el término que engloba a todas aquellas enfermedades causadas por el desarrollo anormal de células. Estas se dividen, crecen y se diseminan sin control en cualquier parte del cuerpo. Diferentes estudios científicos lo han definido y lo han explicado de las siguientes maneras:

“El cáncer se puede originar en cualquier parte del cuerpo. Comienza cuando las células crecen descontroladamente sobrepasando a las células normales, lo cual dificulta que el cuerpo funcione de la manera que debería” (American Cancer Society, 2016).

“Cáncer es el nombre que se da a un conjunto de enfermedades relacionadas. En todos los tipos de cáncer, algunas de las células del cuerpo empiezan a dividirse sin detenerse y se diseminan a los tejidos del alrededor” (Instituto Nacional del Cáncer, 2015).

El cuerpo humano está creado por trillones de células; estas normalmente crecen y se dividen de forma controlada para crear o formar nuevas células a medida que el cuerpo humano lo necesite. Cuando las células se dañan o envejecen son reemplazadas por otras nuevas. Sin embargo, en el cáncer, este proceso se descontrola. Las células crecen sin control y sobrepasan a las células sanas, por lo tanto, las células dañadas o viejas no se mueren y las nuevas se forman sin ser necesario. Estas pueden formar masas de tejido, es decir, un tumor.

Es importante resaltar que no todos los tumores son cancerígenos, por ello se distinguen dos tipos de tumores, los benignos y los malignos. Los benignos, no son tan peligrosos, ya que crecen a un ritmo lento y no se diseminan ni infiltran en los tejidos cercanos. A veces los tumores benignos pueden ser bastante grandes y al extirparse, generalmente no vuelven a crecer. Los tumores malignos se pueden extender e invadir los tejidos cercanos. Cuando crecen, algunas células cancerosas pueden desprenderse y moverse, a través del sistema circulatorio o del linfático, a lugares distantes del cuerpo, es un proceso llamado metástasis. De este modo se forman nuevos tumores lejos del tumor original.

Según la Organización Mundial de la Salud (2018) los cinco tipos de cáncer que causan un mayor número de fallecimientos son: pulmonar (1.69 millones de defunciones), hepático (788.000 defunciones), colorrectal (774.000 defunciones), gástrico (754.000 defunciones) y mamario (571.000 defunciones).

En cuanto a la incidencia del cáncer por grandes zonas geográficas, en 2012, Asia era la región con mayor tasa de incidencia de tumores en todo el mundo, seguido por Europa y Norte América, aunque la tasa de estos últimos (Europa y Norte América) era mucho menor que en la primera (GLOBOCAN, 2012).

En este trabajo vamos a centrarnos, de forma específica en el cáncer de mama. Se va a analizar por un lado su incidencia en la población y, por otro lado, las variables relevantes que afectan o inciden en el tamaño y en el grado del tumor en el momento del diagnóstico.

2. EL CÁNCER DE MAMA/SENO

El cáncer de mama se origina cuando las células en el seno empiezan a crecer de forma descontrolada. Estas células normalmente forman un tumor que a menudo se puede observar en una radiografía o se puede palpar como un bulto. El tumor es maligno si las células crecen penetrando los tejidos circundantes o propagándose a áreas distantes del cuerpo. El cáncer de seno ocurre casi exclusivamente en las mujeres, pero los hombres también pueden padecerlo.

Se distinguen principalmente dos tipos de cánceres de mama; por un lado, el carcinoma ductal, que es el que empieza en las células de los conductos, es decir, se origina en las células que recubren internamente los conductos de la leche y conforman la mayoría de los cánceres de mama; este es el más común. Dentro de este podemos distinguir dos subtipos, carcinoma ductal *in situ* y carcinoma ductal invasivo. El primero es el cáncer que se ubica únicamente en el conducto, y en general se puede curar con tratamiento. Es una afección por la que se encuentran células anormales en el revestimiento de los conductos pero que no se han diseminado fuera del conducto. El segundo, en cambio, es el más grave. El cáncer se ha diseminado fuera del conducto y los tumores se han expandido por todo el cuerpo.

El otro tipo de cáncer es el carcinoma lobular, es el que se origina en los lóbulos. Dentro de este tipo de cáncer encontramos el carcinoma lobular *in situ*, que se ubica solamente en los lóbulos. A pesar de que no se considera exactamente un cáncer, constituye un factor de riesgo para el desarrollo de cáncer de mama invasivo en ambas mamas.

Distintos síntomas pueden llevar a pensar que podemos tener un tumor o un cáncer de mama. Por ejemplo: un bulto en el pecho o en el brazo, cambio de tamaño o forma de la mama, hoyuelos o arrugas en la piel de la mama, pezón hacia dentro del pecho, secreción del pezón que se produce de forma repentina, piel escamosa, enrojecimiento o hinchazón del pecho, pezón o areola, o dolor en la mama.

Para diagnosticar el cáncer de mama se realizan diferentes pruebas. Una de ellas es la mamografía, una radiografía especial de los senos. Si la mamografía muestra algo, el médico puede solicitar más pruebas, como, por ejemplo:

- Ultrasonidos: ondas sonoras de alta energía que rebotan en los órganos y tejidos para crear una imagen.
- Resonancia Magnética: se utiliza un imán, ondas de radio y un ordenador para crear imágenes detalladas.
- Análisis de sangre y orina: se examinan muestras de sangre y orina en busca de riesgos de enfermedad.
- Biopsia: se extraen células o tejidos de la mama y se observan bajo un microscopio. Esto ayuda a detectar células cancerosas.

Diversos estudios (Sociedad Española de Oncología Médica, 2017a; American Cancer Society, 2017a) resaltan que hay varios factores que se deben tener en cuenta para poder prevenir el cáncer de mama: no fumar, limitar el consumo de alcohol, seguir una dieta saludable, mantener un peso acorde con la altura y características de cada uno, hacer ejercicio con regularidad, amamantar a sus bebés si se puede y minimizar la exposición a la radiación en las pruebas de detección si no es médicamente necesario.

Cuando el cáncer se detecta a tiempo aumentan las posibilidades de una recuperación rápida y completa. Una vez el cáncer ha sido detectado, existen diferentes tratamientos. Se utilizan diversas definiciones respecto al tratamiento:

- Tratamiento Local: se refiere al tratamiento dirigido al tumor en su lugar de origen o en alguna localización determinada. Dentro de este tipo de tratamiento se encuentran la cirugía y la radioterapia.
- Tratamiento Sistemático: tratamiento que afecta a todo el organismo. La quimioterapia y la hormonoterapia son tratamientos sistemáticos.
- Tratamiento Adyuvante: se llama así al tratamiento sistemático y/o local administrado tras el primer tratamiento. El objetivo es profiláctico, tanto a nivel sistemático como local, es decir, pretende reducir el riesgo de recidiva del cáncer de mama.
- Tratamiento Neoadyuvante: consiste en administrar un tratamiento sistemático antes de un tratamiento local, con el objetivo de reducir el tamaño del tumor antes de la cirugía.

El avance en la recuperación del cáncer ha sido muy positivo, ya que los diagnósticos que se hacen hoy en día son mucho más precisos y se ajustan más a la necesidad de cada paciente. También hay que tener en cuenta que los resultados en la cura van mejorando a medida que la tecnología se desarrolla con el paso del tiempo. Estas mejoras se pueden deber por ejemplo a los llamados equipos de imagen; estos ayudan a una mayor rapidez en la detección de la enfermedad. Un tratamiento en estadios de la enfermedad más tempranos ayuda a que el cáncer esté menos avanzando y sea más fácil de intervenir.

Además, actualmente se llevan a cabo tratamientos menos agresivos, más ajustados a las necesidades de cada paciente y con menos efectos secundarios. A modo de ejemplo, la resonancia magnética ayuda a los profesionales a detectar mutaciones genéticas que alertan del riesgo de padecer cáncer, por ejemplo, de mama. Por lo tanto, cada paciente tendrá un pronóstico a su medida, teniendo en cuenta que las intervenciones serán lo menos agresivas posibles y se ajustarán las cantidades de radiación a la dimensión exacta del tumor, lo cual ayuda a no intoxicar el resto del cuerpo.

Hoy en día existen muchas expectativas acerca de la cura del cáncer, o por lo menos de la cercanía de esta cura. Son cada día más las investigaciones que se llevan a cabo para encontrar la cura del cáncer; además de grandes donaciones para que estas investigaciones sean llevadas a cabo, con el fin de encontrar un remedio a esta enfermedad que tantas muertes sigue causando.

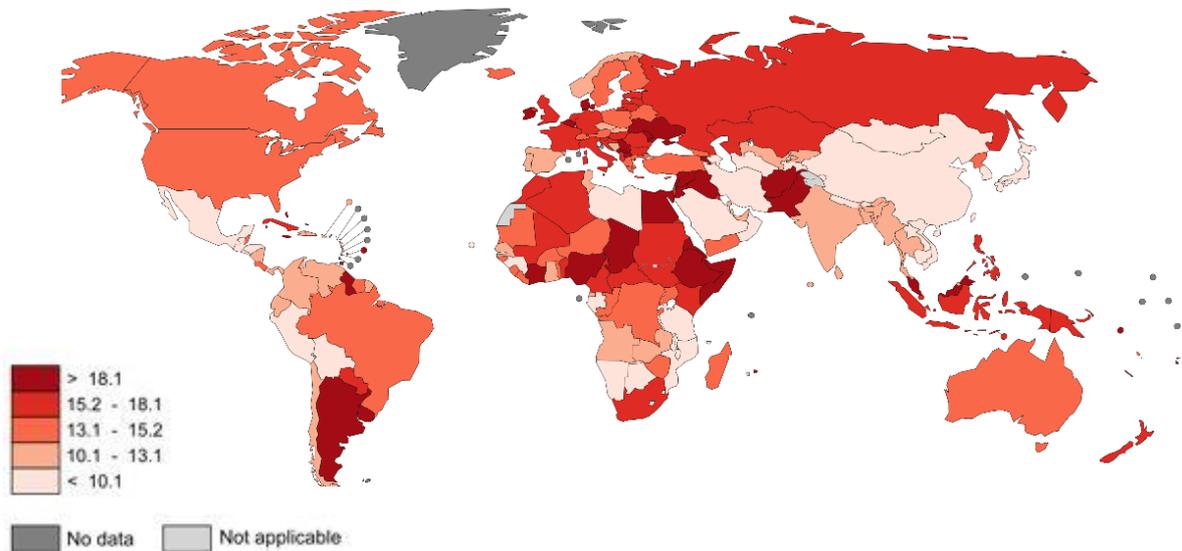
2.1. EL CÁNCER DE MAMA A NIVEL MUNDIAL

El cáncer de seno es el segundo cáncer más común en el mundo (Ferlay et al., 2015; GLOBOCAN, 2012; World Cancer Research Fund, 2015), y es el cáncer más frecuente entre las mujeres, con una estimación de 1.67 millones de nuevos casos de cáncer diagnosticados en 2012 (25% de todos los cánceres), dándose más casos en regiones menos desarrolladas (883.000 casos) que en regiones más desarrolladas (794.000). Las tasas de incidencia varían casi cuatro veces entre algunas regiones del mundo, con tasas que van desde 27 por 100.000 en el África Central y Asia Oriental hasta 92 en América del Norte.

El cáncer de mama es la quinta causa de muerte por cáncer en general (522.000 muertes) y, si bien es la causa más frecuente de muerte por cáncer en mujeres en regiones menos desarrolladas (324.000 muertes, 14,3% del total), ahora es la segunda causa de muerte por cáncer en las regiones más desarrolladas (198.000 muertes, 15,4%), después del cáncer de

pulmón. El rango en las tasas de mortalidad en todo el mundo es menor que la incidencia, debido a la supervivencia más favorable del cáncer de mama en las regiones más desarrolladas, con tasas que varían de 6 por 100.000 en Asia Oriental a 20 por 100.000 en África Occidental.

Figura 1: Tasas de mortalidad del cáncer de mama a nivel mundial



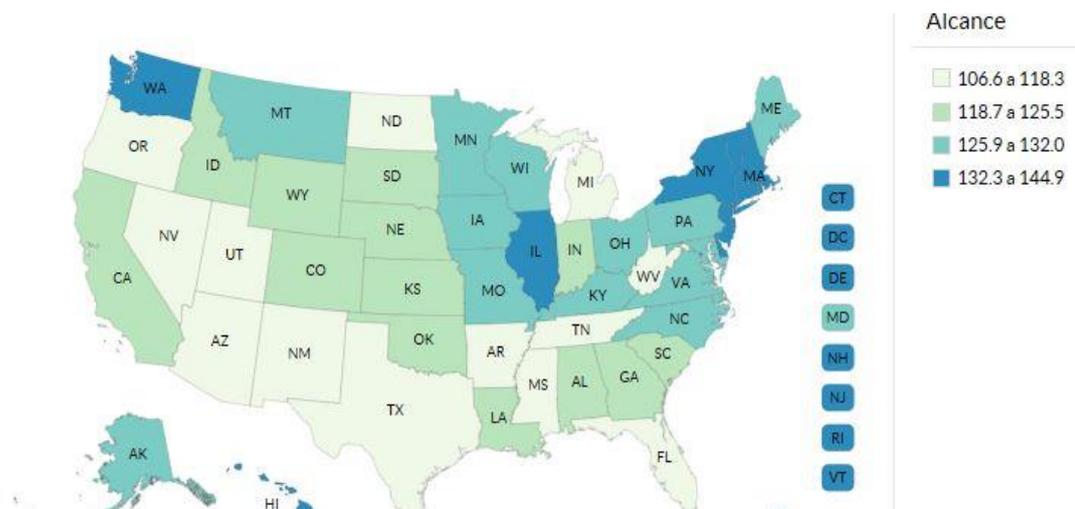
Fuente: GLOBOCAN (2012).

2.2. EL CÁNCER DE MAMA EN ESTADOS UNIDOS

En Estados Unidos el cáncer de mama es el segundo cáncer más común en las mujeres, después del cáncer de piel. Sin embargo, desde 1989, la cantidad de mujeres que mueren por cáncer de mama ha disminuido a un ritmo constante gracias a la detección temprana y a las mejoras en los tratamientos. En la actualidad hay aproximadamente tres millones de mujeres que viven con cáncer de mama en Estados Unidos.

Según los centros para el control y la prevención de las enfermedades, en Estados Unidos, el riesgo de contraer cáncer de mama varía de un estado a otro. Nueva York, Washington, Illinois y Massachussets son los estados con mayor tasa de incidencia de cáncer de mama. Frente a ello, estados como Texas, Nuevo Méjico, Arizona, Nevada y Oregón tienen las menores tasas.

Figura 2: Tasas de incidencia del cáncer de mama en Estados Unidos por estado



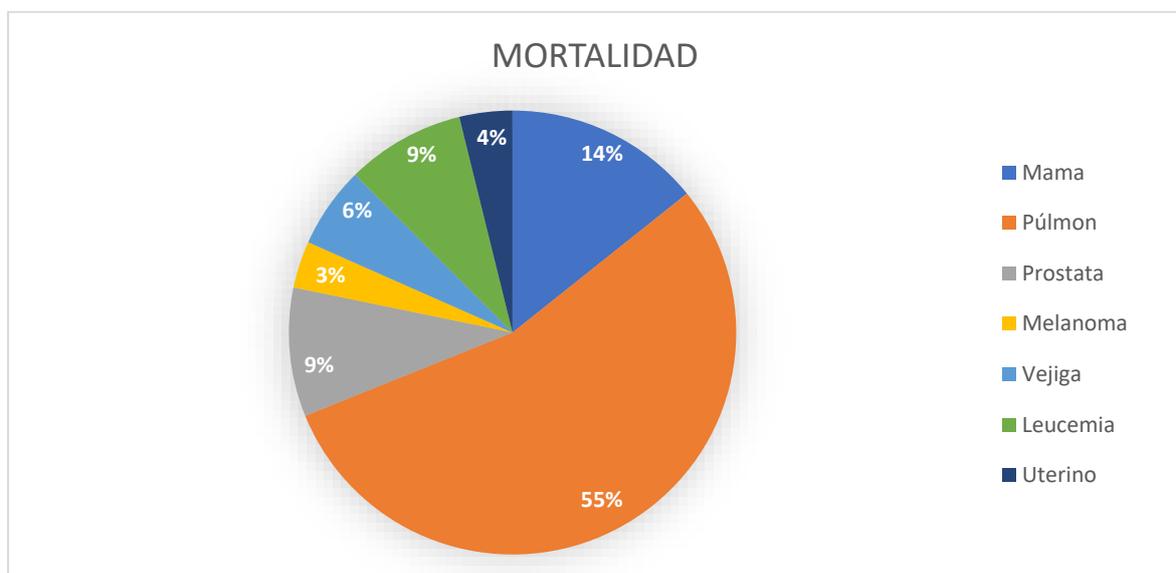
Fuente: Centros para el Control y la Prevención de Enfermedades (2018).

La American Society of Clinical Oncology (2017) concluyó que, en Estados Unidos, en 2017, se habrían diagnosticado 252.710 casos de mujeres con cáncer de mama invasivo y 63.410 mujeres con cáncer de mama *in situ*. En hombres se habrían diagnosticado 2.470 casos entre ambos tipos de cáncer.

Se calcula que este año (2018) se producirán unas 41.070 muertes, 40.610 mujeres y 460 hombres a causa del cáncer de mama.

En el gráfico 1 se observa que, en Estados Unidos, del total de fallecimientos producidos a causa del cáncer, el 14% corresponden al cáncer de mama, solo superado por el cáncer de pulmón con un 55%.

Gráfico 1: Distribución de la mortalidad por tipo de tumor en Estados Unidos

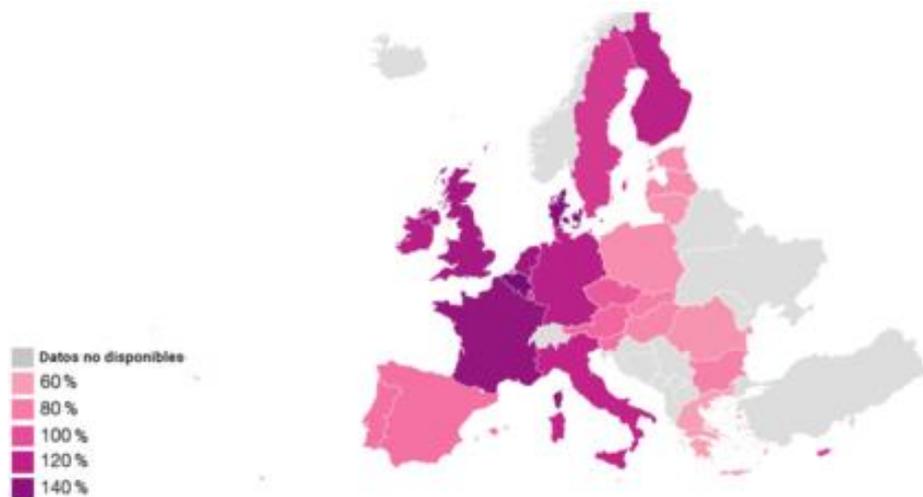


Fuente: Elaboración propia, a partir de American Cancer Society (2017a).

2.3. EL CÁNCER DE MAMA EN ESPAÑA

En España, según la Asociación Española Contra el Cáncer (n.d) y la Red Española de Registros de Cáncer (2016), la incidencia del cáncer es baja; es menor que en Estados Unidos, Canadá y, como podemos ver en la figura 3, también es inferior a la de varios países europeos, tales como Reino Unido, Países Bajos, Bélgica, Alemania, Francia y Suiza.

Figura 3: Tasa de incidencia del cáncer de mama en mujeres en Europa

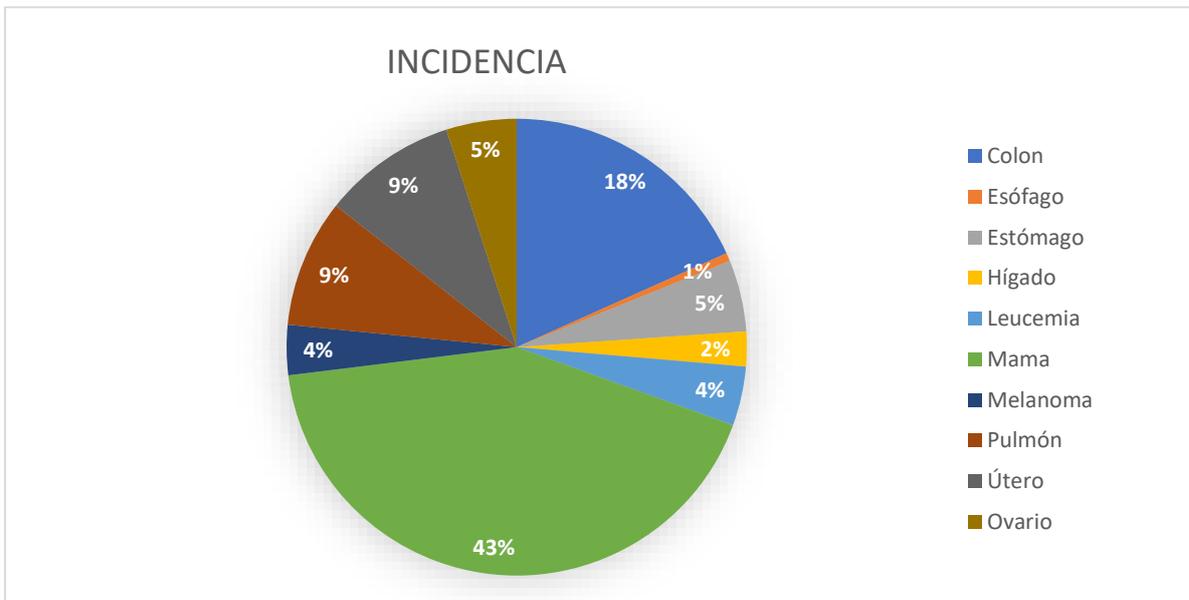


Fuente: Red Española de Registros de Cáncer (REDECAN) (2014).

Se diagnostican alrededor de 25.000 casos de cáncer de mama al año, lo que supone el 43% del total de tumores del sexo femenino (ver gráfico 2). La mayoría de los casos se diagnostican entre los 35 y 80 años, con un máximo entre los 45 y los 65 años, y es el cáncer más común en mujeres en España.

En este país, uno de cada dos hombres y una de cada tres mujeres tendrá un cáncer a lo largo de su vida. Esto se debe, según han explicado los oncólogos españoles, al aumento de la esperanza de vida de la población española, ya que a medida que uno envejece hay más probabilidades de desarrollar cáncer.

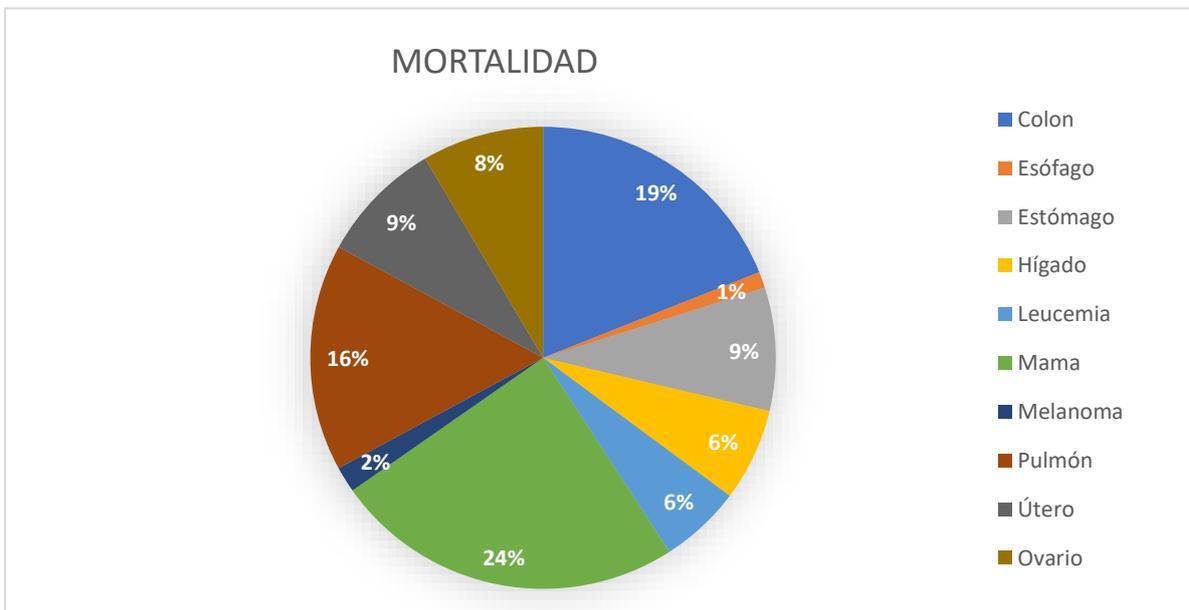
Gráfico 2: Distribución por tipo de tumor del cáncer en España en mujeres



Fuente: Elaboración propia, a partir de REDECAN (2016).

En el gráfico 3 se ve que del total de mujeres que murieron a causa del cáncer en 2014, el 24% fue por causa del cáncer de mama. En España, este tipo de cáncer es el que más muertes provoca entre las mujeres que padecen cáncer (Sociedad Española de Oncología Médica, 2017b).

Gráfico 3: Distribución de la mortalidad por cáncer en función del tipo de tumor

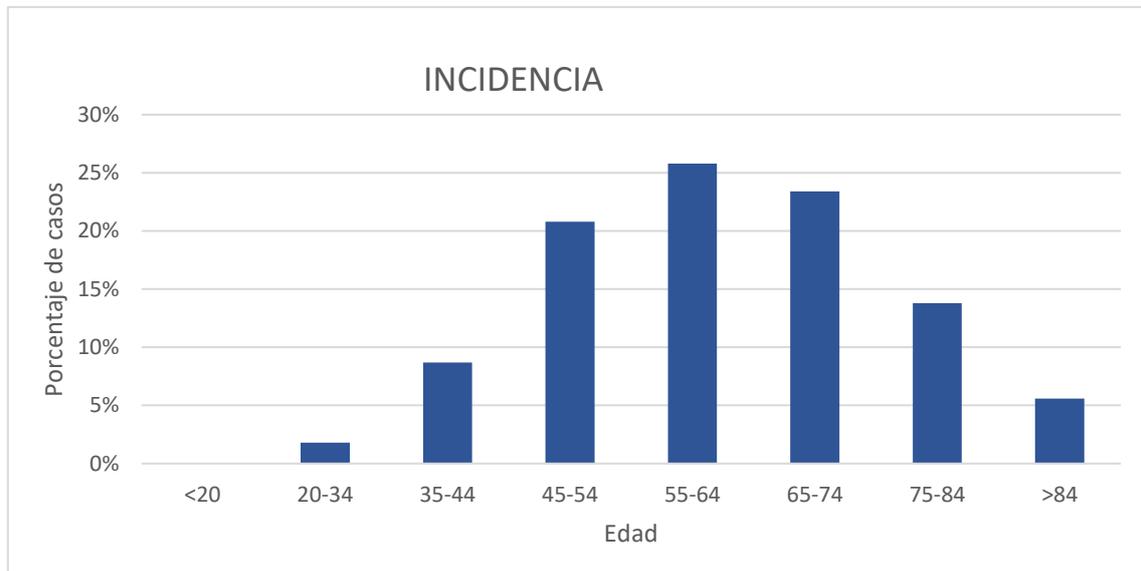


Fuente: Elaboración propia, a partir de SEOM (2017b).

La Sociedad Española de Oncología Médica (2017b) estudió el porcentaje de casos que había de cáncer de mama en las mujeres y la mortalidad de estas según la edad. En el gráfico 4, se observa

que el número de mujeres que padecen cáncer de mama va aumentando con la edad, hasta la franja de 55-64 años. A partir de esta franja de edad este número disminuye.

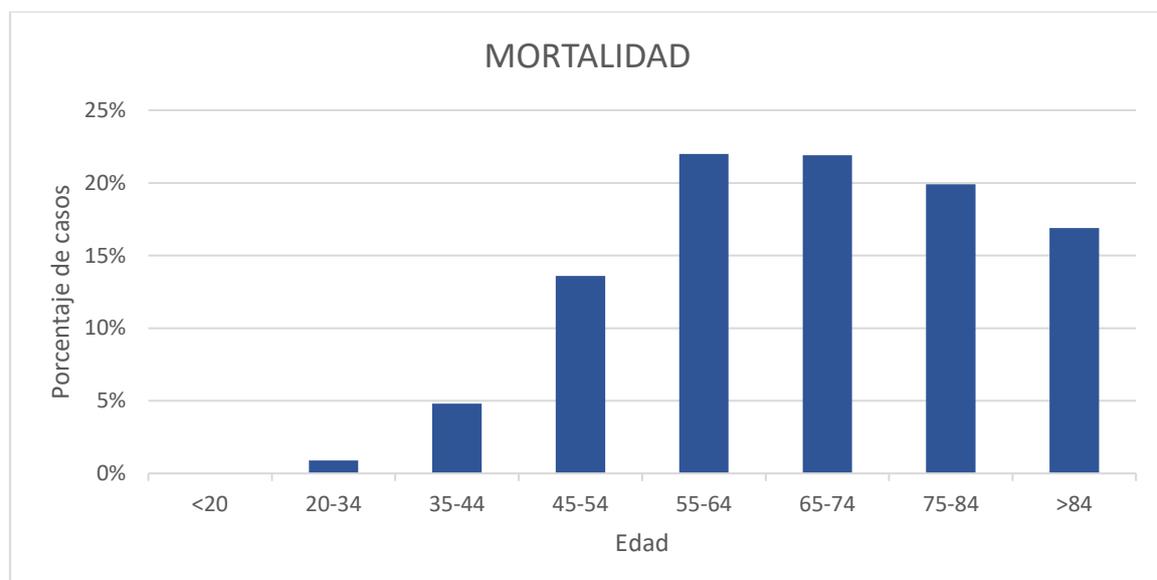
Gráfico 4: Distribución por edad del cáncer de mama en mujeres



Fuente: Elaboración propia, a partir de SEOM (2017b).

En el gráfico 5 podemos ver el número de muertes entre las mujeres diagnosticadas de cáncer de mama. Se observa cómo el porcentaje de mujeres que mueren por el cáncer de mama aumenta con la edad, presentando máximos en las franjas de 55 a 64 y de 65 a 74 años. Posteriormente estos valores disminuyen.

Gráfico 5: Distribución por edad de la mortalidad del cáncer de mama



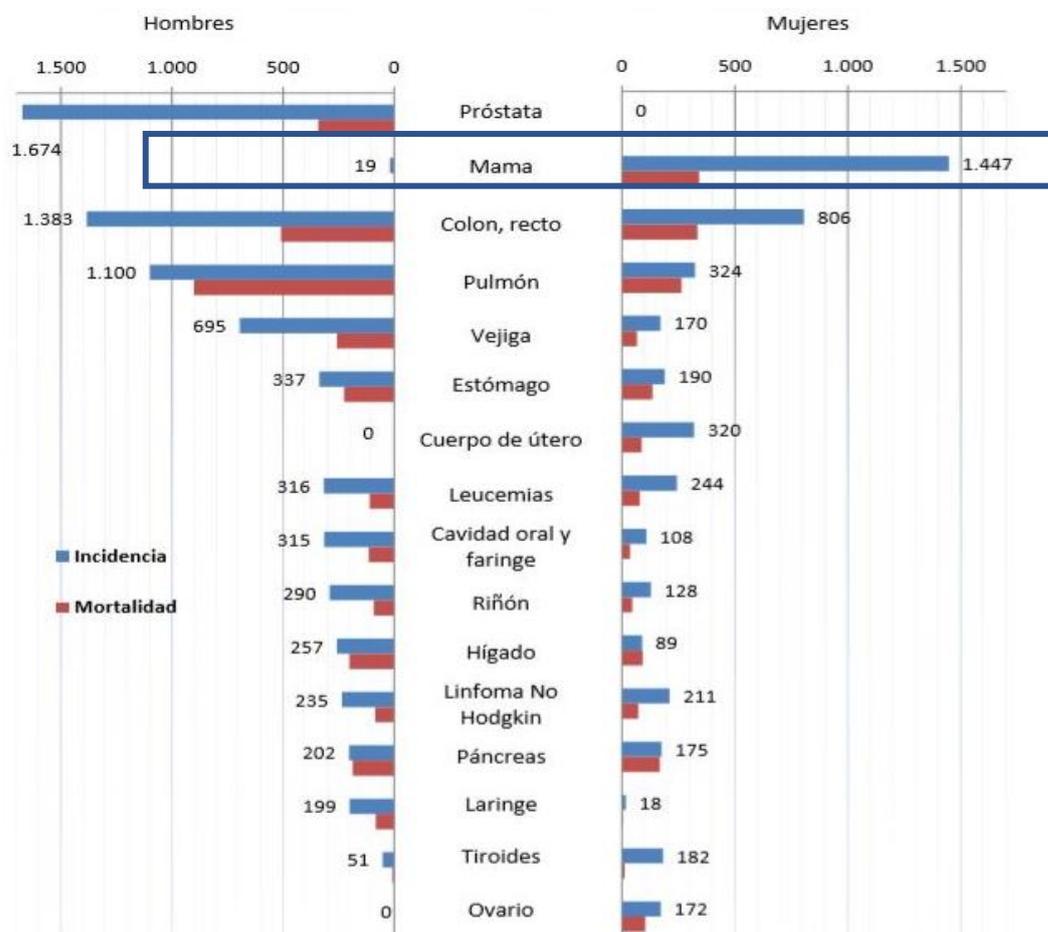
Fuente: Elaboración propia, a partir de SEOM (2017b).

2.4. EL CÁNCER DE MAMA EN EL PAÍS VASCO

En el País Vasco el cáncer de mama es también el más frecuente entre las mujeres (Gobierno Vasco/ Eusko Jaurlaritza, 2010). Entre los años 2000 y 2012, se diagnosticaron 16.362 cánceres de mama invasivos, con una media de 1.260 casos por año (Gobierno Vasco/ Eusko Jaurlaritza, 2018).

En la figura 4 se observa que entre los años 2011 y 2015 se han producido 30.273 defunciones por cáncer en residentes de la Comunidad Autónoma de Euskadi, esto representa una media anual de 6.055 muertes al año (3.739 hombres y 2.316 mujeres). Por sexos, en los hombres el cáncer que más muertes produce es el de pulmón. En las mujeres, el cáncer de mama (14,8%), seguido del de colon-recto (14,4%) y pulmón (11,4%) son los más frecuentes y los que más casos de muertes produce (Gobierno Vasco/ Eusko Jaurlaritza, 2016). Destaca la gran diferencia de casos de cáncer de mama por sexo, en los hombres apenas alcanza el 0,2% y en mujeres el cáncer de mama es el más común. Entre 2009 y 2013 hubo 7.236 casos de cáncer de mama en mujeres, y 94 casos en hombres.

Figura 4: Casos anuales de tumores malignos y defunciones por tipo de tumor y sexo



Fuente: Gobierno Vasco/ Eusko Jaurlaritza (2016).

3. FACTORES SOCIODEMOGRÁFICOS QUE PUEDEN AFECTAR A LA ENFERMEDAD

La revisión de la literatura se ha centrado en diversos artículos científicos de autores españoles, británicos, norteamericanos y latinoamericanos, así como en páginas web de asociaciones y organizaciones que se dedican al estudio, análisis y prevención de la enfermedad.

Existen múltiples variables que afectan al padecimiento del cáncer, bien variables posibles de controlar por el ser humano o variables que no se pueden controlar, como es el caso de las variables sociodemográficas. Para hacer este trabajo se ha decidido considerar, de entre este último tipo de variables, algunas de las más utilizadas en la literatura, como son la raza, el origen hispano, la edad en el momento del diagnóstico y el estado civil.

- Diversos autores (Bradley et al., 2001; Wingo et al., 1999; Shinagawa, 2000; Lannin et al., 1998) concluyen que hay mucha disparidad en el cáncer de mama dependiendo de la **raza** de la mujer. Las mujeres afroamericanas tienen más probabilidad que las mujeres blancas de tener cáncer de mama en una etapa tardía en el momento del diagnóstico, y, en consecuencia, de tener resultados menos favorables a su curación. También hay evidencia sobre las diferencias que hay en el diagnóstico, el tratamiento y la supervivencia del cáncer de mama entre las mujeres de diferente raza o de diferentes situaciones económicas. Sin embargo, es difícil para los gobiernos y grupos de investigación dirigir intervenciones que reducirían la brecha en la atención médica entre estos dos grupos de mujeres.

Algunos autores, como por ejemplo American Society of Clinical Oncology (2017), indican que las mujeres de raza blanca tienen más probabilidades de desarrollar cáncer de mama que las mujeres de raza negra a partir de los 45 años, pero entre las mujeres menores de 45 años las mujeres de raza negra tienen más probabilidades de padecer la enfermedad que las mujeres jóvenes de raza blanca. Las mujeres de raza negra también tienen más probabilidades de morir a causa de la enfermedad.

Por otro lado, en los trabajos de American Cancer Society (2017b) y del Instituto Nacional de Cáncer (2017a) también se incluye el **origen hispano** o no de la mujer como factor explicativo tanto de su supervivencia como del pronóstico de la enfermedad.

Cabe señalar que persisten también algunas disparidades raciales en la supervivencia del cáncer de mama, incluso cuando se controla el estado socioeconómico en los análisis. Sin embargo, el efecto general de la raza sobre el diagnóstico del cáncer y la

supervivencia se reduce enormemente cuando los modelos analíticos incluyen el ingreso como variable explicativa. Aunque en Estados Unidos la raza se considera con frecuencia como la mayor causa de desigualdad, la discapacidad, la ubicación geográfica, los ingresos y la educación también se incluyen en el mismo grupo (Instituto Nacional del Cáncer, 2018).

- En relación al nivel de ingresos, según el artículo de Serral et al. (2018), si bien las diferencias encontradas respecto de la evolución del cáncer de mama han ido disminuyendo a lo largo del tiempo, estas siguen persistiendo, y esto se debe a que son las mujeres de las clases sociales más bajas las que se realizan un menor control mamográfico.

En el caso de Estados Unidos, Mickey et al. (1995) y Garbers et al. (2003) concluyen que las mujeres con bajos ingresos que residen en áreas metropolitanas podrían ser menos propensas a someterse a exámenes de detección de cáncer de seno que las mujeres más pudientes que residen en las mismas áreas. Además, las mujeres con educación menor a la secundaria tienen menos probabilidades de realizarse exámenes de detección de cáncer de mama que las mujeres con más educación, especialmente en áreas donde una mayor proporción de mujeres tienen estudios.

- Es indiscutible que la **edad** es uno de los indicadores de salud más universal. Como señalan Perry et al. (2008), y siguiendo las recomendaciones de la European Breast Cancer Network (EBCN), en España se viene desarrollando desde mediados de los años noventa del siglo pasado, un programa de detección precoz de cáncer, en el que se invita, por medio de una carta, a mujeres de edades de entre 50 y 69 años a realizarse una mamografía cada dos años (Castells et al., 2009). Es un programa que se instauró en todas las comunidades autónomas en 2006.

Según el artículo de Serral et al. (2018), el 91,9% de las mujeres se hacen una mamografía una vez en la vida (9 de cada 10 mujeres). De las mujeres que se hacen un control mamográfico, alrededor del 83% se lo realiza con una frecuencia menor o igual a dos años. Son las mujeres de mayor edad (60-69 años) las que muestran una mayor proporción de controles mamográficos, y el motivo suele ser por un programa preventivo.

En general, se puede concluir que son las mujeres de menor edad a las que con más frecuencia se les detecta el cáncer de mama en un grado más avanzado de la

enfermedad. En cambio, cuando el tumor está en un grado inicial, es a las mujeres con mayor edad a las que se les detecta en una mayor proporción.

- El **estado civil** también tiene efecto sobre el tiempo de supervivencia (Sánchez, 2012). Según este autor, vivir en pareja y el cariño de la misma aumenta las posibilidades de recuperación ya que se han comprobado sus propiedades terapéuticas, las cuales ayudan a afrontar la situación.

Parece ser que tener unas sólidas relaciones afectivas y formar parte activa de un grupo social que incluya amigos, familiares y pareja puede disminuir la mortalidad en algunos casos de cáncer de mama. Y es que no es lo mismo afrontar un problema de salud, como el cáncer, en soledad, que sentir el apoyo de los seres queridos (Sánchez, 2012).

En el estudio internacional llevado a cabo entre 1990 y 2010 en el que participaron el Hospital Universitario Vall d'Hebron y el Vall d'Hebron Instituto de Oncología (VHIO), y presentado en el 50 Congreso de la Sociedad Americana de Oncología (ASCO), se concluye que el estado civil puede influir en el pronóstico de la enfermedad; mientras que las mujeres casadas tienen un mejor pronóstico del cáncer, con una supervivencia del 89% a los 5 años, en el colectivo de pacientes no casadas (en las que se incluyen aquellas que nunca se han casado, separadas y viudas), la supervivencia es del 82%. En el caso de las mujeres casadas, el cáncer de mama se detecta en una fase inicial y los tumores son más pequeños (ABC Salud, 2014).

Los resultados del estudio también mostraron que la mejora de la supervivencia entre las mujeres casadas es mayor cuanto más avanzado está el cáncer de mama. Así, entre las pacientes en las fases I, II, III y IV del cáncer de mama se observa una mejora significativa a los cinco años de un 1%, 3%, 9% y 7%, respectivamente. Por otro lado, parece que también podría existir una relación entre el estado civil y la edad en la que se diagnostica la patología: en el caso de las mujeres no casadas la supervivencia empeora con el paso de los años.

En el caso de Estados Unidos, en general, el 78,5% de las mujeres mayores de 40 años afirman haberse hecho una mamografía con una frecuencia inferior a dos años. Entre las mujeres que declararon ingresos familiares anuales inferiores a 15.000\$, el 68,4% se hizo una mamografía antes de que pasen dos años entre una y otra; el 75,3% de las mujeres con ingresos familiares de entre 15.000\$ y 34.999\$ y el 82,5% de las mujeres con ingresos familiares superiores a 50.000\$ se habían realizado una mamografía en los últimos dos años (CDC, 2005).

4. ANÁLISIS ESTADÍSTICO Y ECONOMETRICO DE CASOS DE CÁNCER DE MAMA

Como se ha descrito en los apartados anteriores, el cáncer de mama está causado por muchas razones. Hay muchas variables que hacen que el riesgo de padecer esta enfermedad sea mayor o menor. Algunas de las causas dependen de uno mismo, como por ejemplo el consumo de tabaco o de alcohol; pero hay otras causas que están fuera del control del ser humano, por ejemplo, la genética, o determinadas variables sociodemográficas.

En este estudio vamos a analizar la relevancia de algunas de dichas variables que afectan al diagnóstico y a la evolución del cáncer de mama en las mujeres, pero que no se pueden controlar por uno mismo. En concreto, y siguiendo los trabajos de la Asociación Española Contra el Cáncer (2017), del Instituto Nacional del Cáncer (2017b) y de la American Cancer Society (2017c), las variables que se van a considerar son: el estado civil, la raza, el origen hispano, y la edad en el momento del diagnóstico.

Los datos para hacer este análisis provienen de SEER CANCER RESEARCH DATABASE (2017) del Instituto Nacional de Cáncer de Estados Unidos. En esta base de datos se recogen casos de mujeres de EE. UU. con cáncer de mama diagnosticadas entre los años 1990 y 2014. Todas las mujeres de la muestra padecen esta enfermedad. Por ello vamos a analizar el efecto de las variables consideradas sobre la detección precoz de cáncer considerando como indicadores, el grado y el tamaño en el momento del diagnóstico.

En este estudio únicamente hemos tenido en cuenta a personas del sexo femenino que han sido diagnosticadas de cáncer de mama en el año 2014.

El objetivo de este trabajo es analizar si existe alguna relación entre determinadas variables sociodemográficas tales como el estado civil, la raza, el origen hispano y la edad con el grado y el tamaño del tumor en el momento del diagnóstico. También se ha considerado la posición del tumor (lateralidad).

Para ello, se ha realizado en primer lugar el análisis descriptivo de las mismas. Posteriormente, se han realizado contrastes de independencia entre el grado del tumor y las variables sociodemográficas consideradas en el estudio. Además, se han estimado distintos Modelos de Regresión Lineal para determinar la relación entre el tamaño del tumor en el momento del diagnóstico y las diferentes variables sociodemográficas consideradas, utilizando el método de Mínimos Cuadrados Ordinarios (MCO).

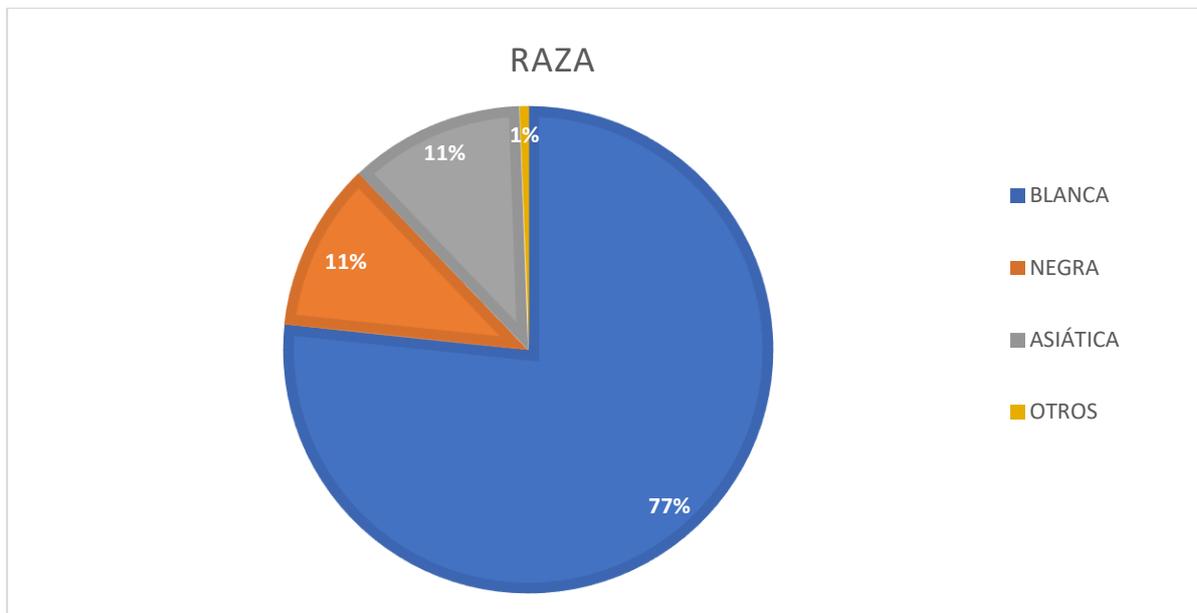
Para describir estadísticamente las variables utilizadas en el estudio y estimar los modelos especificados se ha utilizado el programa econométrico Gretl (Versión 1.9.2) y para hacer los contrastes de independencia se ha utilizado el programa SPSS (Versión 23.0.0.0).

4.1. DESCRIPCIÓN DE LAS VARIABLES

Para poder mostrar con más claridad la distribución de las variables analizadas (raza, estado civil, edad y grado) en la muestra, se realizan gráficos sectoriales.

En el gráfico 6 se observa que, entre todas las mujeres, la gran mayoría es de raza blanca, aproximadamente un 77%, mientras que las de raza negra y raza asiática representan un 11% cada una.

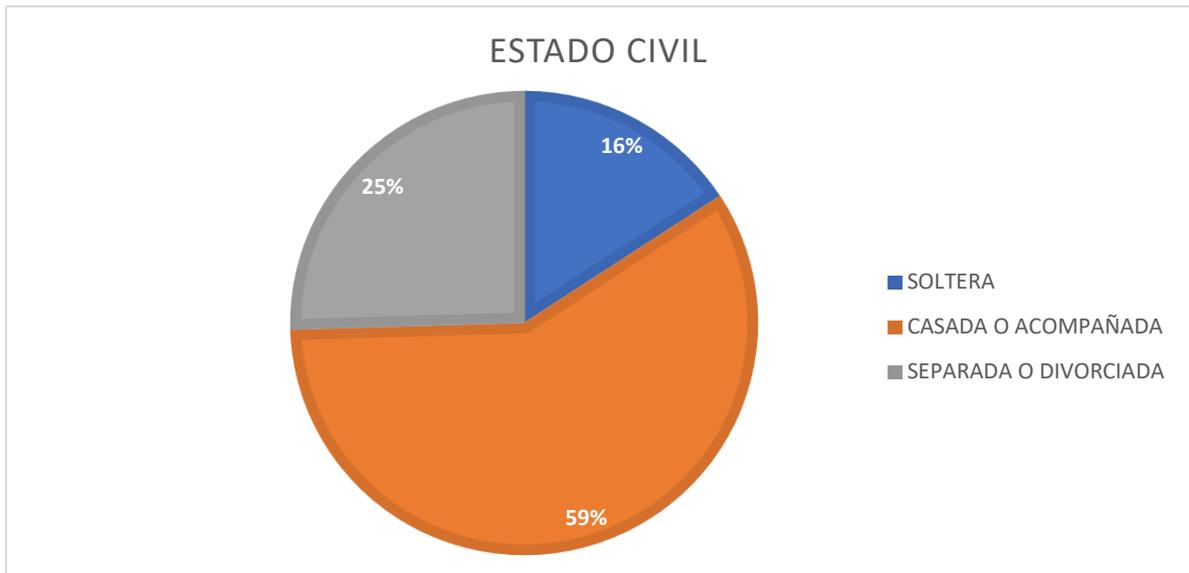
Gráfico 6: Distribución por raza de las mujeres con cáncer de mama



Fuente: *Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).*

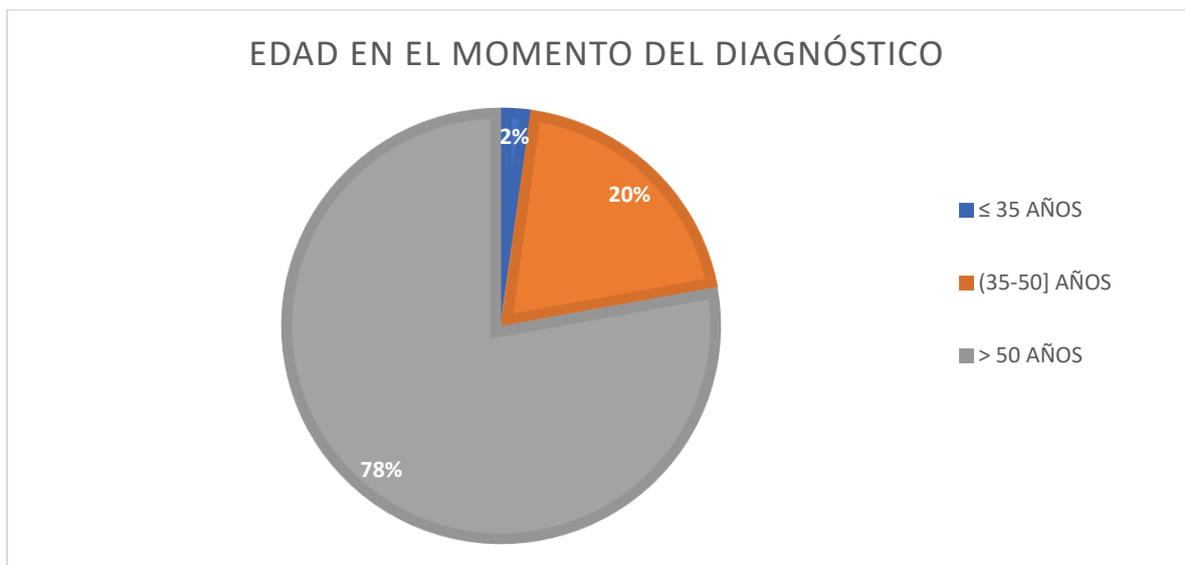
En el gráfico 7 se ve cuál es la distribución de las mujeres según el estado civil, es decir, si las mujeres están casadas o acompañadas, solteras o separadas o divorciadas. Comprobamos que la gran mayoría de las mujeres en este caso están casadas o acompañadas, estas son el 59% del total, seguidas por las separadas o divorciadas con un 25%.

Gráfico 7: Distribución por estado civil de las mujeres con cáncer de mama



Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Gráfico 8: Distribución por edad de las mujeres con cáncer de mama

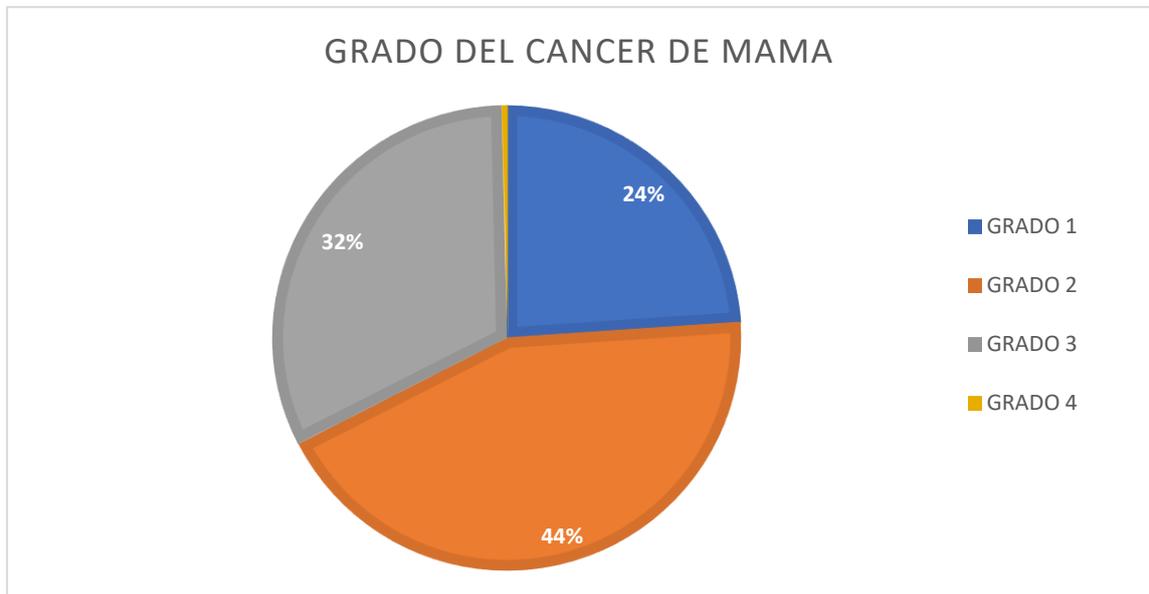


Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

En el gráfico 8 se refleja la distribución la edad de las mujeres en el momento del diagnóstico. Se han agrupado a las mujeres en tres grupos de edad, menor o igual a 35 años, entre 35 y 50 años y, por último, mujeres mayores de 50 años. Se ve cómo la gran mayoría son del grupo de edad de mayores de 50 años con un 78%. La minoría es la de edad menor o igual de 35 años.

En el gráfico 9 vemos cuál es el grado del tumor en el momento que se detecta el cáncer de mama, siendo lo más frecuente el segundo grado (44%) y lo minoritario, casi residual, el cuarto grado (1%).

Gráfico 9: Distribución por grado de las mujeres con cáncer de mama



Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Además de estas cuatro variables, también se recogen las variables origen hispano y lateralidad. Respecto a la primera, la gran mayoría de las mujeres de la muestra es de origen no hispano, ya que son el 92,81%, frente a la proporción de mujeres de origen hispano, que es únicamente un 7,19%.

En lo que se refiere a la lateralidad, la diferencia de que el cáncer se diagnostique en la mama derecha o en la izquierda es mínima; la mama derecha es la de mayor frecuencia (51%).

4.2. CONTRASTES DE INDEPENDENCIA

En este apartado se realizan las pruebas de independencia entre el grado del tumor y cada una de las variables socioeconómicas consideradas.

Cuando se realiza un contraste de independencia, la muestra está clasificada en categorías de acuerdo a dos variables, X e Y , cuya independencia se quiere contrastar. La primera variable está clasificada de acuerdo a k' categorías y la segunda a k'' categorías.

Si las variables son independientes, la probabilidad conjunta de cada cruce de categorías i, j , será igual al producto de las probabilidades marginales, es decir, para todo $i = 1, \dots, k'$, $j = 1, \dots, k''$,

$$P_{ij} = P_{i\cdot}P_{\cdot j} \quad (1)$$

donde,

P_{ij} = probabilidad conjunta de la categoría i, j .

$P_{i\cdot}$ = Probabilidad marginal de la categoría i -ésima de la variable X .

$P_{\cdot j}$ = Probabilidad marginal de la categoría j -ésima de la variable Y .

Por lo tanto, las hipótesis nula y alternativa de este contraste se pueden expresar como:

$$H_0: P_{ij} = P_{i\cdot}P_{\cdot j} \quad \text{para todo } i, j$$

$$H_1: P_{ij} \neq P_{i\cdot}P_{\cdot j} \quad \text{para algún } i, j$$

Para cada categoría i, j , se conocen las frecuencias observadas en la muestra, n_{ij} , y se calculan las frecuencias estimadas bajo el supuesto de que la hipótesis nula es cierta, \hat{E}_{ij} , de la siguiente forma:

$$\hat{E}_{ij} = n\hat{P}_{ij} = n\hat{P}_{i\cdot}\hat{P}_{\cdot j} = n \frac{n_{i\cdot}}{n} \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n} \quad (2)$$

donde,

$n_{i\cdot}$ = la frecuencia marginal de la variable X para la categoría i -ésima.

$n_{\cdot j}$ = la frecuencia marginal de la variable Y para la categoría j -ésima.

n = número total de elementos de la muestra.

El estadístico de contraste, Z , recoge las desviaciones existentes para cada categoría i, j , entre ambos tipos de frecuencias, las observadas, n_{ij} , y las estimadas bajo el supuesto de que la hipótesis nula es cierta, \hat{E}_{ij} . Se define como,

$$Z = \sum_{i=1}^{k'} \sum_{j=1}^{k''} \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \quad (3)$$

Este estadístico, bajo la hipótesis nula, sigue asintóticamente una distribución $\chi^2_{(k'-1)(k''-1)}$.

La regla de decisión será rechazar la hipótesis nula de independencia al nivel de significación α si el valor muestral del estadístico es superior al valor crítico, es decir,

$$z > \chi^2_{(k'-1)(k''-1), \alpha} \quad (4)$$

que es equivalente a comprobar si:

$$p - \text{valor} < \alpha \quad (5)$$

A continuación, se exponen los resultados obtenidos para cada una de las variables. En todos los contrastes que se realizan se utiliza el nivel de significación del 5%.

- **RAZA**

Tabla 1: Relación raza y grado del tumor

	GRADO 1	GRADO 2	GRADO 3	GRADO 4	Total
RAZA 1 Blanca	5118 (25,3%)	8861 (43,9%)	6151 (30,4%)	74 (0,4%)	20204 (100%)
RAZA 2 Negra	500 (17,1%)	1184 (40,5%)	1234 (42,2%)	7 (0,2%)	2925 (100%)
RAZA 3 Asiática	658 (21,4%)	1367 (44,4%)	1042 (33,8%)	14 (0,5%)	3081 (100%)
RAZA 4 Otros	27 (18,2%)	69 (46,6%)	51 (34,5%)	1 (0,7%)	148 (100%)
Total	6303 (23,9%)	11481 (43,6%)	8478 (32,2%)	96 (0,4%)	26531 (100%)

Valor Estadístico: 207,012; gl: 9; p-valor: <0,0001.

Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Dado que el p-valor es inferior a 0,05, se rechaza la hipótesis nula de que la raza y el grado de cáncer de mama son independientes entre sí. Por ello, consideramos que esta variable sí es relevante en la determinación del grado.

En la tabla 1 se puede observar que cuando el cáncer no está muy avanzado (grados 1 y 2), son las mujeres de raza blanca a las que con mayor frecuencia se les detecta esta enfermedad. En cambio, a medida que avanza la enfermedad, es decir, a medida que el grado aumenta, las mujeres de raza negra son a las que en mayor proporción se les localiza el tumor; esto se aprecia sobre todo en el grado 3.

Por lo tanto, las mujeres de raza blanca van a tener mejor pronóstico, comparando especialmente con las mujeres de raza negra, que son las de peor pronóstico.

- **ORIGEN HISPANO**

Tabla 2: Relación origen hispano y grado del tumor

	GRADO 1	GRADO 2	GRADO 3	GRADO 4	Total
NO HISPANO	5879 (23,9%)	10749 (43,8%)	7854 (32,0%)	87 (0,4%)	24569 (100%)
HISPANO	456 (24,0%)	774 (40,8%)	660 (34,8%)	9 (0,5%)	1899 (100%)
Total	6335 (23,9%)	11523 (43,5%)	8514 (32,2%)	96 (0,4%)	26468 (100%)

Valor Estadístico: 8,587; gl: 3; p-valor: 0,035.

Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Dado el p-valor obtenido, también en este caso se rechaza la hipótesis nula de que el origen y el grado del cáncer son independientes entre sí. Por ello, consideramos que esta variable sí es relevante en la determinación del grado en el momento del diagnóstico.

En la tabla 2 se muestra la diferencia que hay entre las mujeres de origen no hispano e hispano en el momento de detectar esta enfermedad. Se puede observar que en el grado 1 la diferencia es mínima, obteniéndose las diferencias más claras en los grados 2 y 3. En el grado 2, a las mujeres de origen no hispano se les detecta con mayor frecuencia el cáncer, pero, en cambio, a medida que la enfermedad avanza la proporción cambia, ya que son las mujeres hispanas a las que con mayor frecuencia se les detecta el tumor.

Cabe destacar que es una variable significativa al nivel de significación del 5% pero no lo sería a niveles inferiores (por ejemplo, 1%). En este caso, el valor del estadístico no nos lleva a resultados tan concluyentes como otras variables de las consideradas, que resultan significativas a niveles de significación inferiores.

Además, el ser o no de origen hispano está interactuando con la raza de la mujer, ya que tanto entre las mujeres de origen hispano como entre las no hispanas puede haber mujeres de raza blanca, negra, asiática u otros. Las diferencias significativas se obtienen claramente entre mujeres blancas (hispanas y no hispanas) pero no sucede lo mismo en las otras razas.

- **EDAD**

Tabla 3: Relación edad y grado del tumor

	GRADO 1	GRADO 2	GRADO 3	GRADO 4	Total
EDAD 1 ≤35	38 (6,4%)	206 (34,6%)	347 (58,3%)	4 (0,7%)	595 (100%)
EDAD 2 (35-50]	1042 (19,9%)	2143 (40,9%)	2029 (38,8%)	22 (0,4%)	5236 (100%)
EDAD 3 >50	5255 (25,5%)	9174 (44,5%)	6138 (29,7%)	70 (0,3%)	20637 (100%)
Total	6335 (23,9%)	11523 (43,5%)	8514 (32,2%)	96 (0,4%)	26468 (100%)

Valor Estadístico: 392,491; gl: 6; p-valor: <0,0001.

Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Se observa cómo también en este caso se rechaza la hipótesis nula de que la edad y el grado de cáncer de mama son independientes entre sí. Por ello, consideramos que esta variable sí es relevante en la determinación del grado.

Podemos comprobar que son las mujeres mayores de 50 años a las que con mayor frecuencia se les detecta el cáncer de mama cuando este está en un grado inicial y la paciente tiene más posibilidades de curarse. Pero una vez que la enfermedad está en un grado superior, son las mujeres más jóvenes, las de edad menor o igual a 35 años, a las que con mayor frecuencia se les detecta el cáncer de mama.

- **ESTADO CIVIL**

Tabla 4: Relación estado civil y grado del tumor

	GRADO 1	GRADO 2	GRADO 3	GRADO 4	Total
ESTADO CIVIL 1 Soltera	857 (21,8%)	1640 (41,7%)	1420 (36,1%)	18 (0,5%)	3935 (100%)
ESTADO CIVIL 2 Casada/ Acompañada	3567 (24,0%)	6495 (43,6%)	4775 (32,1%)	48 (0,3%)	14885 (100%)
ESTADO CIVIL 3 Divorciada/ Separada	1579 (24,9%)	2812 (44,3%)	1925 (30,4%)	26 (0,4%)	6342 (100%)
Total	6026 (23,8%)	11027 (43,5%)	8177 (32,3%)	93 (0,4%)	25162 (100%)

Valor Estadístico: 41,320; gl: 6; p-valor: <0,0001.

Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Para esta variable, al igual que sucedía con las anteriores, se rechaza la hipótesis nula de que el estado civil y el grado de cáncer de mama son independientes entre sí. Por lo tanto, consideramos que sí es relevante en la detección temprana del cáncer.

Al analizar esta tabla, nos damos cuenta de que cuando una mujer está soltera, la proporción en la que se le detecta el cáncer de mama en grado 1 es inferior a las mujeres que tienen otro estado civil.

Además, cabe destacar que la edad está interactuando con el estado civil de la mujer, ya que entre las mujeres más jóvenes hay mayor proporción de solteras y entre las de mayor edad, hay más proporción de casadas o en pareja, divorciadas y separadas. Por último, reseñamos, que no sabemos con seguridad si las mujeres que entran dentro del grupo de divorciadas o separadas se han vuelto a casar o emparejar con otra persona.

- **LATERALIDAD**

Tabla 5: Relación lateralidad y grado del tumor

	GRADO 1	GRADO 2	GRADO 3	GRADO 4	Total
LATERALIDAD 1 Derecha	3156 (24,1%)	5728 (43,7%)	4170 (31,8%)	44 (0,3%)	13098 (100%)
LATERALIDAD 2 Izquierda	3179 (23,8%)	5794 (43,4%)	4339 (32,5%)	52 (0,4%)	13364 (100%)
Total	6335 (23,9%)	11522 (43,5%)	8509 (32,2%)	96 (0,4%)	26462 (100%)

Valor Estadístico: 1,811; gl: 3; p-valor: 0,613.

Fuente: Elaboración propia, a partir de SEER CANCER RESEARCH DATABASE (2017).

Por último, cuando se analiza la lateralidad, no se encuentran diferencias significativas, por lo que consideramos que no hay ninguna relevancia de la variable “lateralidad” sobre el grado del tumor en el momento del diagnóstico.

Además, debemos indicar que esta no es una variable sociodemográfica, pero puede ser interesante analizar si existen diferencias, como señalan algunos autores. En este sentido, la Universidad Nacional del Nordeste (2000) señala que el cáncer de mama afecta a la mama izquierda con una frecuencia un poco mayor que a la derecha.

• INTERPRETACIÓN DE RESULTADOS

Haciendo referencia a la literatura y teniendo en cuenta los resultados obtenidos en los contrastes de independencia, se puede concluir que son las mujeres de raza blanca las que mejor pronóstico tienen y las mujeres de raza negra las que peor. Respecto a ser o no ser de origen hispano cabe señalar que son las mujeres de origen hispano a las que con mayor frecuencia se les detecta la enfermedad en un estado avanzado. Hay que tener en cuenta, como se ha comentado anteriormente, que dentro del origen hispano o no hispano puede haber mujeres de cualquier raza (blanca, negra, asiática, u otros).

Por otro lado, son las mujeres de edad avanzada a las que primero se les detecta el tumor cuando este está en el grado inicial. Sin embargo, a las mujeres más jóvenes se les detecta el cáncer en estado más avanzado.

En cuanto al estado civil de la mujer, se puede concluir que a las mujeres casadas o acompañadas se les detecta en mayor proporción cuando la enfermedad está en un grado inicial. Al igual que ocurre entre la raza y el origen hispano, el estado civil y la edad también están interactuando entre sí ya que, dentro de las mujeres solteras, la mayoría son las más jóvenes, y entre las mujeres acompañadas o casadas suelen estar las de mayor edad.

4.3. ANÁLISIS DE REGRESIÓN

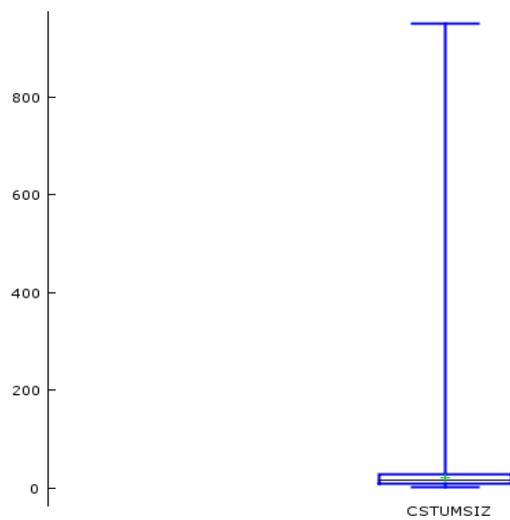
En este apartado, y teniendo en cuenta los resultados de los contrastes de independencia que se acaban de presentar, se realiza un análisis de regresión del tamaño del tumor sobre las variables anteriores. Esto se hace porque creemos que hay relación entre el tamaño y el grado del cáncer y, por lo tanto, es de esperar que las variables que afectan a uno también afecten al otro.

Para el análisis se van a redefinir las variables usadas anteriormente, porque para realizar el análisis de regresión y las diferentes estimaciones, las variables edad, grado raza y estado civil se van a tratar como variables ficticias. Antes de realizar el análisis vamos a fijarnos en las características de la variable a explicar, es decir, el tamaño del tumor.

Se observa que el valor máximo de dicha variable es 988mm. Tumores de este tamaño parecen ser excesivamente grandes para un cáncer de mama y, además, se puede apreciar que a partir de un determinado valor el número de casos es excesivamente pequeño en relación a la mayoría del colectivo. Por ello, se ha realizado un diagrama de caja del tamaño del tumor (figuras 5 y 6), para analizar la dispersión de la variable y poder determinar cuál es el límite a partir del cual los

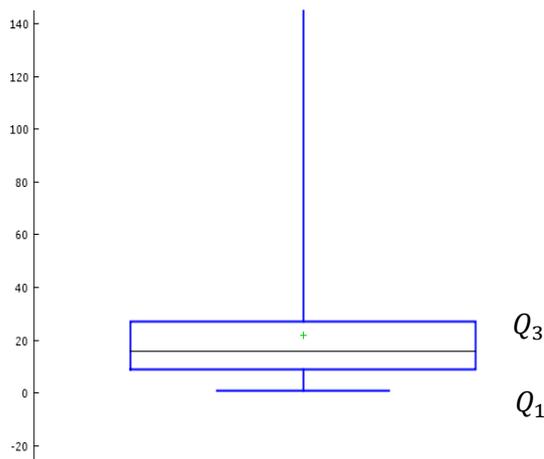
valores observados se consideran atípicos. Por esta razón, la muestra se restringe a aquellas mujeres para las que la variable tamaño del tumor está por debajo del valor que se obtiene al sumar al tercer cuartil 1,5 veces el recorrido intercuartílico ($q_3 + 1,5R_{13}$), que en este caso es 54 milímetros. Por lo tanto, se realizarán las regresiones únicamente para aquellos casos en que el tamaño del tumor esté entre 0 y 54 milímetros, ya que en este intervalo es donde mayor proporción de cánceres de mama hay. Una vez restringida la muestra, el tamaño medio del tumor es de aproximadamente 17,5mm.

Figura 5: Diagrama de caja del tamaño del tumor



Fuente: Elaboración propia.

Figura 6: Diagrama del tamaño del cáncer de mama. (Zoom de figura 5)



MÁXIMO	950	-
MEDIANA= Q_2	16	+
MEDIA	21,952	-----
MINIMO	1	-
Q_1	9	—
Q_3	27	—

Fuente: Elaboración propia.

Las variables implicadas se resumen en la tabla 6. Los valores típicos que aparecen recogidos son los que corresponden a la muestra ya restringida a que el tamaño del tumor sea inferior o igual que 54 milímetros.

Podemos observar en esta tabla que el 59,4% de las mujeres de nuestro estudio están casadas, el 76,9% de ellas son de raza blanca, un 7% es de origen hispano y un 78,7% tienen más de 50 años. En cuanto al grado del cáncer que padece cada mujer, el grado 4, que es el más peligroso, lo tiene solo el 0,2% del total de mujeres. El de grado 1, el menos peligroso, es el padecido por un 25% de las mujeres. El de grado 2, poco peligroso, lo padecen el 43% de las mujeres, y finalmente, el de grado 3, lo padece el 31% del total de mujeres.

Tabla 6: Variables del modelo

VARIABLE FICTICIA	NOTACIÓN	DESCRIPCIÓN	MEDIA	DESVIACIÓN TIÍPICA
Tamaño del cáncer	CSTUMSIZ	Variable dependiente. Medida en milímetros. (0,54].	17,553	11,582
Estado civil	EST. CIVIL 1	Toma valor 1 si la mujer está soltera; 0 en el resto de los casos.	0,1543	0,3612
	EST. CIVIL 2	Toma valor 1 si la mujer está casada o acompañada; 0 en el resto de los casos.	0,5939	0,4911
	EST. CIVIL 3	Toma valor 1 si la mujer se ha divorciado o separado.; 0 en el resto de los casos.	0,2518	0,4341
Raza	RAZA 1	Toma valor 1 si la mujer es blanca; 0 en el resto de los casos.	0,7691	0,4214
	RAZA 2	Toma valor 1 si la mujer es negra; 0 en el resto de los casos.	0,1074	0,3097
	RAZA 3	Toma valor 1 si la mujer es de raza asiática; 0 en el resto de los casos.	0,1178	0,3224
	RAZA 4	Toma valor 1 si la mujer es de otra raza; 0 en el resto de los casos.	0,0056	0,0745
Origen hispano	HISPANO	Toma valor 1 si la mujer es de origen hispano; 0 en el resto de los casos.	0,0702	0,2555

Edad en el momento del diagnóstico	EDAD 1	Toma valor 1 si la edad de la paciente en el momento del diagnóstico es menor o igual de 35 años; 0 en el caso contrario.	0,0198	0,1394
	EDAD 2	Toma valor 1 si la edad > que 35 años y ≤ que 50 años; 0 en el caso contrario.	0,1932	0,3948
	EDAD 3	Toma valor 1 si la edad es > que 50 años; 0 en el caso contrario.	0,7870	0,4095
	EDAD	Variable cuantitativa. Medida en años.	61,499	13,031
Grado de cáncer	GRADO 1	Toma valor 1 si el cáncer de la mujer es de grado 1; 0 en el resto de los casos.	0,2501	0,4331
	GRADO 2	Toma valor 1 si el cáncer es de grado 2; 0 en el resto de los casos.	0,4369	0,4960
	GRADO 3	Toma valor 1 si el cáncer es de grado 3; 0 en el resto de los casos.	0,3104	0,4626
	GRADO 4	Toma valor 1 si el cáncer es de grado 4; 0 en el resto de los casos.	0,0025	0,0507
Lateralidad	LATERALIDAD	Toma valor 0 si es el lado derecho; 1 en el lado izquierdo.	0,5051	0,5000

Fuente: Elaboración propia.

- **ESTIMACIÓN POR MCO**

A la vista de la revisión de la literatura realizada y de los resultados obtenidos anteriormente, el modelo general que se propone estimar es:

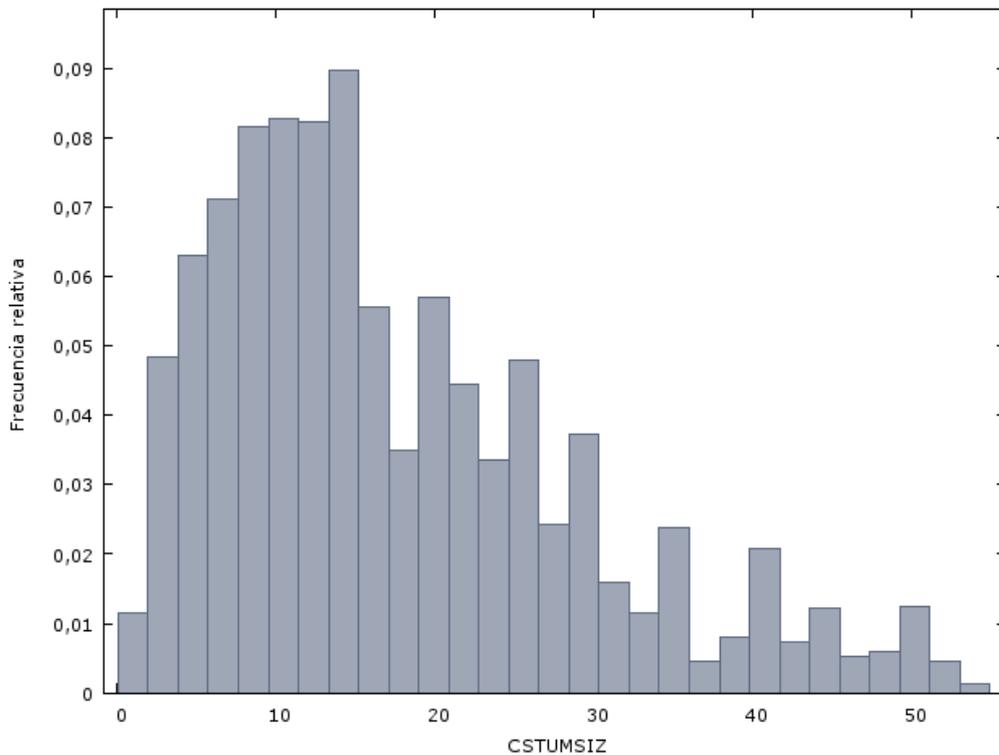
$$\text{Tamaño tumor} = f(\text{edad, raza, estado civil, origen hispano, lateralidad, grado})$$

en el que se va a incluir en la especificación la edad (edad a la que la mujer es diagnosticada de cáncer de mama) tanto como variable continua (MODELO 1) como en intervalos (≤ 35 años), (35-50] años y (> 50 años) (MODELO 3). También se considera la posibilidad de que la influencia de la edad no sea lineal y por ello se introduce también como variable explicativa la *edad*² (MODELO 2).

Por último, y teniendo en cuenta el histograma de la variable dependiente (figura 7), también se ha estimado un modelo logarítmico lineal (MODELOS 4 y 5).

La estimación de los distintos modelos considerados aparece en la tabla 7.

Figura 7: Histograma del tamaño del cáncer de mama



Fuente: Elaboración propia.

Para poder elegir la especificación más adecuada del modelo podemos considerar los contrastes de significación, tanto individuales como conjuntos.

○ **Contraste de significación individual**

Las hipótesis nula y alternativa para el contraste de significación individual son:

$$H_0: \beta_i = 0 \quad i = 2, \dots, k.$$

$$H_1: \beta_i \neq 0$$

El estadístico del contraste es:

$$t = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t_{T-K}$$

Tabla 7: Estimación por MCO

	M1	M2	M3	M4	M5
EST. CIVIL 2	-1,11481*** (7,2410)	-0,82590*** (5,3469)	-1,12106*** (7,3927)	-0,06066*** (5,9194)	-0,06206*** (6,1199)
RAZA 1	-0,80984*** (4,5323)	-0,85130*** (4,7868)	-0,74391*** (4,1870)	-0,02987** (2,5028)	-0,02579** (2,1709)
HISPANO	1,19391*** (4,0520)	1,13907*** (3,8844)	1,07775*** (3,6709)	0,06601*** (3,3529)	0,05924*** (3,0174)
GRADO 1	-8,28705*** (41,3204)	-8,14237*** (40,7468)	-8,14655*** (40,8307)	-0,54365*** (40,5329)	-0,53547*** (40,1361)
GRADO 2	-3,81460*** (21,8902)	-3,72781*** (21,4836)	-3,71423*** (21,4137)	-0,23614*** (20,2707)	-0,23027*** (19,8537)
LATERALIDAD	-0,04011 (0,2716)	-0,04168 (0,2835)	-0,03391 (0,2300)	0,00276 (0,2799)	0,00310 (0,3146)
EDAD	-0,01708*** (2,8807)	-0,64711*** (14,9949)	-	-	-
EDAD ²	-	0,00514*** (14,7375)	-	-	-
EDAD 1 ≤35	-	-	4,21510*** (7,9104)	-	0,22898*** (6,4264)
EDAD 2 (35-50]	-	-	1,24805*** (6,6061)	-	0,06246*** (4,9445)
LogEDAD	-	-	-	-0,04083* (-1,7771)	-
R ²	0,07929	0,08817	0,08297	0,07407	0,07643

Nota: ***, **, * representan los niveles de significación del 1%, 5%, y 10%, respectivamente. Los estadísticos t están entre paréntesis.

Fuente: Elaboración propia.

Si el valor muestral del estadístico t es tal que $|t| < t_{T-K, \frac{\alpha}{2}}$, no rechazamos la H_0 para un nivel de significación α , es decir, la variable es estadísticamente no significativa.

Como se puede observar en la tabla 7, todos los estadísticos t muestrales mostrados en los diferentes modelos son mayores que $t_{(24419-9)\frac{0,05}{2}} = 1,96$ por lo que se rechazan las hipótesis nulas correspondientes a todos los parámetros, es decir, las variables son estadísticamente relevantes de forma individual.

Esto no ocurre en el caso de la variable lateralidad. En ninguna de las especificaciones consideradas el estadístico muestral es superior a 1,96, por lo que la variable lateralidad no es estadísticamente significativa, de forma individual. Este resultado es coincidente con el obtenido en el contraste de independencia.

○ **Contraste de significación conjunta**

Las hipótesis nula y alternativa para el contraste de k-1 restricciones lineales son:

$$H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1: \text{algún } \beta_i \neq 0$$

Y el estadístico del contraste:

$$F = \frac{R^2/(k-1)}{1 - R^2/(T-K)} \sim F_{K-1, T-K}$$

Si el valor muestral del estadístico F, es tal que $F < F_{K-1, T-K, \alpha}$ no rechazamos la H_0 para un nivel de significación α , con lo que se puede concluir que las variables no son conjuntamente significativas.

En los casos considerados, y como ejemplo para el M2:

$$F_2 = \frac{0.088166/(9-1)}{1 - 0.088166/(24419-9)} = 295.0279409$$

$$F_2 = 295.0279409 > F_{(8, 24419), 0.05} = 1,93879$$

Como el valor muestral del estadístico F, es tal que $F > F_{(K-1, T-K), 0.05}$, rechazamos la hipótesis nula para un nivel de significación $\alpha=0,05$, es decir, las variables consideradas son conjuntamente significativas. A modo de resumen, se puede decir que todas las variables consideradas, excepto la lateralidad, son significativas individual y conjuntamente.

- **CONTRASTES DE HETEROCEDASTICIDAD**

En los datos de sección cruzada tales como los utilizados en este trabajo es frecuente encontrar problemas de heterocedasticidad, aunque no de autocorrelación (más frecuente en datos de serie temporal), por ello a continuación se procede a estudiar este problema utilizando los contrastes de White y de Breusch- Pagan.

En presencia de heterocedasticidad, el estimador Mínimo Cuadrático Ordinario es lineal, insesgado y consistente, pero no es de varianza mínima. Su matriz de varianzas y covarianzas se define como:

$$V(\hat{\beta}_{MCO}) = \sigma^2 (X'X)^{-1} X' \Omega X (X'X)^{-1} \quad (6)$$

En tanto que el estimador de la matriz de varianzas y covarianzas del estimador MCO cuando hay homocedasticidad es:

$$\hat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2 (X'X)^{-1} \quad (7)$$

Utilizar este último estimador para hacer inferencia cuando hay heterocedasticidad no es adecuado. Los estadísticos t y F habituales para hacer inferencia sobre β , definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$, son inapropiados ya que es un estimador sesgado e inconsistente.

La dificultad que entraña el conocimiento de Ω , hace interesante poder contar con un estimador de la matriz $V(\hat{\beta}_{MCO})$ consistente y robusto a la posible existencia de heterocedasticidad y, de esta forma, derivar estadísticos válidos, al menos asintóticamente, para contrastar hipótesis sobre el vector de coeficientes β utilizando $\hat{\beta}_{MCO}$.

Si existe homocedasticidad, la ecuación (7) proporciona un estimador consistente de $V(\hat{\beta}_{MCO})$ mientras que si existe heterocedasticidad no lo hará.

Para formular la mayor parte de los contrastes posibles de heterocedasticidad, es necesario especificar, al menos de manera aproximada, la naturaleza de la heterocedasticidad. Sería deseable poder contrastar una hipótesis de la forma

$$H_0: \sigma_i^2 = \sigma^2, \forall i$$

$$H_1: \sigma_i^2 \neq \sigma^2 \text{ para algún } i$$

Sin embargo, se presenta la dificultad de la estimación en un modelo con n parámetros desconocidos (las n varianzas).

○ **Contraste de White**

Si no hay heterocedasticidad, (6) dará un estimador consistente de la matriz de varianzas y covarianzas del estimador de MCO, mientras que, si existe, no lo hará. White (1980) ha propuesto un contraste estadístico basado en la estimación de una regresión auxiliar de los errores MCO al cuadrado sobre las variables explicativas de la regresión inicial. El estadístico (nR^2) se distribuye asintóticamente como una chi-cuadrado con $P-1$ grados de libertad, donde P es el número de regresores en la regresión auxiliar (sin incluir el término constante).

Tabla 8: Contrastes de heterocedasticidad

	M1	M2	M3	M4	M5
WHITE	400,96391	399,75948	377,29915	112,64851	88,89777
BREUSCH Y PAGAN	458,86453	483,27714	465,24405	88,47427	72,68247

Fuente: Elaboración propia.

En la tabla 8 aparecen el valor de los estadísticos para los cinco modelos considerados. En todos los casos $nR^2 > \chi_{p-1,0.05}^2$ por lo que se rechaza la hipótesis nula para un nivel de significación $\alpha=0.05$, y por lo tanto existe heterocedasticidad.

○ **Contraste de Breusch y Pagan**

Breusch y Pagan (1979) proponen un contraste de heterocedasticidad donde la hipótesis alternativa es bastante general:

$$H_0: E(u_i^2) = \sigma^2, \quad \forall_i$$

$$H_1: \sigma^2 = \sigma^2 g(\alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi})$$

Las variables Z_{pi} pueden o no ser variables explicativas del modelo de interés; en cualquier caso, deben ser observadas. Si todos los coeficientes de la combinación lineal $Z_i^t \alpha$ fuesen 0 excepto α_0 , la varianza sería homocedástica, $\sigma_i^2 = \sigma^2 g(\alpha_0)$. Por tanto, un contraste de la hipótesis nula de homocedasticidad vendría dado por la siguiente hipótesis:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

donde se contrastan p restricciones lineales. El contraste consiste en realizar la regresión auxiliar de los residuos normalizados,

$$\hat{e}_i^2 = \frac{\hat{u}_{MCO,i}^2}{\hat{u}'\hat{u}/N} \quad i = 1, 2, \dots, N$$

sobre las variables Z , y calcular el estadístico:

$$BP = \frac{SCE}{2} \xrightarrow{d, H_0} X_p^2$$

siendo p los grados de libertad (igual al número de variables Z_{pi}). Rechazamos la H_0 de homocedasticidad a un nivel de significación α , si el valor muestral del estadístico excede del valor crítico, esto es,

$$BP > X_{p,\alpha}^2.$$

La interpretación del contraste es la siguiente: si las perturbaciones fuesen homocedásticas, las variables Z_{pi} no deberían tener poder explicativo acerca de los residuos transformados y por tanto la SCE debería ser pequeña. Si SCE/2 es grande, rechazaremos la H_0 y concluiríamos en que existe heterocedasticidad.

En este caso, para los cinco modelos de la tabla 7, $BP > \chi_{p,0.05}^2$ por lo que se rechaza la H_0 para un nivel de significación $\alpha=0.05$, y existe heterocedasticidad.

- **ESTIMACIÓN POR MCO CON DESVIACIONES TÍPICAS ROBUSTAS**

A la vista de los contrastes de heterocedasticidad mostrados en la tabla 8, se puede concluir que es posible que exista un problema de heterocedasticidad. Por ello se reestiman los modelos propuestos utilizando el estimador robusto de la matriz de varianzas y covarianzas propuesta por White.

En la tabla 9 aparecen recogidos los resultados de tales modelos. Como se puede observar, los principales resultados de la tabla 7 sobre significación individual y conjunta se mantienen.

Tabla 9: Estimación por MCO. Matriz de varianzas y covarianzas robustas

	M1	M2	M3	M4	M5
ESTADO CIVIL 2	-1,11481*** (7,2155)	-0,82590*** (5,3238)	-1,12106*** (7,3433)	-0,06066*** (5,9380)	-0,06205*** (6,1344)
RAZA 1	-0,80983*** (4,3994)	-0,85129*** (4,6396)	-0,74390*** (4,0583)	-0,02987** (2,4441)	-0,02579** (2,1178)
HISPANO	1,19391*** (3,8965)	1,13907*** (3,7226)	1,07775*** (3,5249)	0,06601*** (3,3234)	0,05924*** (2,9867)
GRADO 1	-8,28705*** (43,2004)	-8,14237*** (42,5626)	-8,14655*** (42,6213)	-0,54365*** (40,7692)	-0,53547*** (40,4206)
GRADO 2	-3,81460*** (20,6130)	-3,72781*** (20,2169)	-3,71423*** (20,1598)	-0,23614*** (20,4617)	-0,23027*** (20,0689)
LATERALIDAD	-0,04011 (0,2716)	-0,04168 (0,2835)	-0,03391 (0,2300)	0,00276 (0,2799)	0,00311 (0,3147)
EDAD	-0,01708*** (2,7736)	-0,64711*** (14,0267)	-	-	-
EDAD ²	-	0,00514*** (13,8458)	-	-	-
EDAD 1 ≤35	-	-	4,21510*** (7,0588)	-	0,22898*** (6,9498)
EDAD 2 (35-50]	-	-	1,24805*** (6,4250)	-	0,06246*** (4,8593)
LogEDAD	-	-	-	-0,04083* (-1,8120)	-
R ²	0,07929	0,08817	0,08297	0,07407	0,07643

Nota: ***, **, * representan los niveles de significación del 1%, 5%, y 10%, respectivamente. Los estadísticos t están entre paréntesis.

Fuente: Elaboración propia.

• INTERPRETACIÓN DE LOS PARÁMETROS

De los resultados obtenidos de la estimación de los diferentes modelos propuestos se puede concluir que:

- Apenas hay diferencias en los resultados de las diferentes especificaciones consideradas.
- Todas las variables incluidas, excepto la lateralidad, son estadísticamente significativas de forma individual.

Teniendo en cuenta que tanto la significación individual como conjunta en todos los casos es similar, el modelo elegido, M3, es aquel en el que las variables (tamaño del tumor y edad en el momento del diagnóstico) intervienen de forma lineal y, en el que la edad se incluye en intervalos, ya que esto nos permite identificar el grupo de edad en el que el tamaño del tumor es mayor.

En cuanto a la interpretación de los coeficientes, los signos obtenidos concuerdan, en general, con la expectativa que se tenía a priori derivada de los trabajos previos. En concreto, en relación al Modelo 3:

- **Estado civil.** Que la mujer esté casada o acompañada hace disminuir el tamaño del tumor en 1,12 mm aproximadamente, *ceteris paribus*.
- **Raza.** Que la mujer sea de raza blanca hace disminuir el tamaño del tumor en 0,74 mm aproximadamente, *ceteris paribus*.
- **Origen hispano.** Que la mujer sea de origen hispano hace aumentar el tamaño del tumor en 1,08 mm aproximadamente, *ceteris paribus*.
- **Grado del tumor.** Que la mujer tenga un tumor en las fases iniciales hace disminuir el tamaño del mismo en 8,14 mm aproximadamente, cuando está en la fase inicial respecto a cuando está en la fase última, y en 3,71 de la fase intermedia a la fase final, *ceteris paribus*.
- **Edad.** Al incluir la edad como variables ficticias, ser menor de 35 años, hace aumentar el tamaño del tumor en 4,21 mm, respecto de las mujeres de más de 50 años. Aquellas mujeres en la edad intermedia tienen de media un tumor de 1,25mm más que las mujeres de más edad.

Por último, decir que, la figura 7 (pág. 35) muestra una mayor frecuencia observada en aquellos valores (para el tamaño del tumor) múltiplos de 5 y mayores de 15 milímetros. Sin tener seguridad creemos que se debe a la aproximación en la medición del tamaño. Si este es el

problema nos encontramos con que la variable dependiente de los distintos modelos está medida con error. Tal y como se señala en Stock-Watson (2012) el error de medida en Y aumenta la varianza de la regresión y de los parámetros estimados, pero no induce sesgo, es decir los parámetros estimados siguen siendo insesgados.

5. CONCLUSIONES

El avance en la recuperación de los enfermos de cáncer ha sido muy positivo en los últimos años, gracias a los avances que se han ido haciendo tanto en el ámbito de la salud como en el de la tecnología. En este último, los equipos de imagen ayudan a una mayor rapidez en la detección de la enfermedad. Por ello, los análisis y los diagnósticos que se hacen a los pacientes son mucho más precisos y se ajustan más a las necesidades de cada uno. Otro elemento importante a tener en cuenta es que actualmente, los tratamientos son menos agresivos y tienen menos efectos secundarios.

Algunas de las variables más utilizadas por los autores (Bradley et al., 2001; Perry et al., 2008; Castell et al.; 2009 y Serral et al., 2018) en los estudios relacionados con la incidencia o con la evolución del cáncer de mama son la raza, el origen hispano, la edad y el estado civil, variables que han sido utilizadas en este trabajo.

A partir de los datos proporcionados por SEER CANCER RESEARCH DATABASE (2017) se han realizado contrastes de independencia para determinar si las variables consideradas son relevantes sobre el grado del cáncer en el momento del diagnóstico, llegándose en todos los casos a la conclusión de que sí tienen efecto sobre dicho grado.

Basándonos en tales resultados, se ha realizado un análisis de regresión para explicar el tamaño del tumor de mama que tienen las mujeres americanas diagnosticadas en el año 2014. Se ha llegado a la conclusión de que todas las variables consideradas (excepto la lateralidad) son estadísticamente significativas de forma individual. Lo mismo ocurre con la significación conjunta. Del análisis realizado se puede concluir que ser mujer casada o acompañada (respecto a soltera, separada o divorciada), ser de raza blanca (respecto a las de raza negra, asiática y otras) y tener tumor de grado 1 y 2 (respecto de los tumores de grado 3 o 4) hacen disminuir, en media, el tamaño del tumor, mientras que ser mujer menor de 35 años y tener entre 35 y 50 años lo hace aumentar, cuando se les compara con el grupo de las mayores de 50 años, que ha sido aquel cuya variable ficticia se ha omitido.

Resaltar que el ajuste del modelo es relativamente bajo en todas las especificaciones consideradas (alrededor del 8%), y del 8,30% en la especificación elegida. Sin embargo, hay que tener en cuenta que, además de las variables sociodemográficas consideradas, son de mucha importancia las variables médicas, no consideradas en este trabajo al no disponer de información.

6. BIBLIOGRAFÍA

- ABC Salud (2014, 27 de mayo). Las mujeres con pareja tienen mejor pronóstico y sobreviven más al cáncer de mama.
- American Cancer Society (2016). ¿Qué es el cáncer? Disponible en <https://www.cancer.org/es/cancer/aspectos-basicos-sobre-el-cancer/que-es-el-cancer.html>
- American Cancer Society (2017a). Cancer facts & Figures. Disponible en <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf>
- American Cancer Society (2017b). Cáncer de seno. Disponible en <https://www.cancer.org/es/cancer/cancer-de-seno/acerca/que-es-el-cancer-de-seno.html>
- American Cancer Society (2017c). ¿Cómo se origina el cáncer de seno? Disponible en <https://www.cancer.org/es/cancer/cancer-de-seno/acerca/como-se-forma-el-cancer-de-seno.html>
- American Society of Clinical Oncology (ASCO) (2017). Cáncer de mama. Disponible en <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama>
- Asociación Española Contra el Cáncer (2017). Cáncer de mama. Disponible en <https://www.aecc.es/SobreElCancer/CancerPorLocalizacion/CancerMama/Paginas/cancredemama.aspx>
- Asociación Española Contra el Cáncer (n.d). Evolución del cáncer de mama. Disponible en <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-mama/evolucion-cancer-mama>
- Borrás, J.M. (2015). La estrategia contra el cáncer del Sistema Nacional de Salud. Gaceta Medica de Bilbao. Vol. 112. Número 1. (pp.40).
- Bradley, C.J.; Given, C.W.; Roberts C. (2001). Disparities in cancer diagnosis and survival. Cancer. Vol. 91. Número 1. (pp.178-188).

- Breusch, T. S., Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*. (pp.1287-1294).
- Cabasés, J.M. (2015). Economía de la salud y cáncer. *Gaceta Médica de Bilbao*. Vol. 112. Número 1. (pp.41).
- Castells, X.; Sala, M.; Salas, D.; Asuncion, N.; Zubizarreta, R. & Casamitjana, M. (2009). Reflexiones sobre las prácticas de diagnóstico precoz del cáncer en España. *Gaceta Sanitaria*. Vol. 23. Número 3.
- CDC (2005). Breast Cancer Screening and Socioeconomic Status - 35 Metropolitan Areas, 2000 and 2002. Disponible en <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5439a2.htm>
- Centro para el Control y la Prevención de las Enfermedades (2018). Tasas de cáncer de mama por estado. Disponible en <https://www.cdc.gov/spanish/cancer/breast/statistics/state.htm>
- Ferlay J.; Soerjomataram I.; Dikshit R.; Eser S.; Mathers C.; Rebelo M.; Parkin D.M.; Forman D.; Bray F. (2015). Cancer incidence and mortality worldwide. *International Journal of Cancer*. *International Journal of Cancer*. Vol. 136. Issue 5. (pp.E359-E386).
- Garbers, S.; Jessop, D.J.; Foti, H.; Uribelarrea, M.; Chiasson M.A. (2003). Barriers to breast cancer screening for low-income Mexican and Dominican women in New York City. *Journal of Urban Health*. Vol. 80. Issue 1. (pp.81–91).
- GLOBOCAN (2012). Breast Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012. Disponible en <http://globocan.iarc.fr/old/FactSheets/cancers/breast-new.asp>
- Gobierno Vasco/ Eusko Jaurlaritza (2010). El cáncer en el País Vasco. Disponible en http://www.euskadi.eus/contenidos/informacion/registros_epidem_presentacion/es_def/adjuntos/cancer.pdf

- Gobierno Vasco/ Eusko Jaurlaritza (2016). Informe 2016 Salud pública y adicciones. Disponible en http://www.euskadi.eus/contenidos/informacion/informes_salud_publica/es_def/adjuntos/informe-SPyA-2016.pdf
- Gobierno Vasco/ Eusko Jaurlaritza (2018). Supervivencia de cáncer en la comunidad autónoma vasca. Disponible en http://www.euskadi.eus/contenidos/informacion/publicaciones_departamento/es_def/adjuntos/osagin/estudio_superviv_cancer_es.pdf
- Greene, W. (2012). Econometric Analysis. Ed Pearson.
- Instituto Nacional del Cáncer (2015). ¿Qué es el cáncer? Disponible en <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>
- Instituto Nacional del Cáncer (2017a). Genética del cáncer. Disponible en <https://www.cancer.gov/espanol/cancer/causas-prevencion/genetica>
- Instituto Nacional del Cáncer (2017b). Cáncer de seno (mama)- Versión para pacientes. Disponible en <https://www.cancer.gov/espanol/tipos/seno>
- Instituto Nacional del Cáncer (2018). Desigualdades por cáncer. Disponible en <https://www.cancer.gov/espanol/cancer/naturaleza/desigualdades>
- Lannin, D.R.; Matthews, H.F.; Mitchell, J. Swanson, M.S.; Swanson, F.H. & Edwards, M.S. (1998). Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. JAMA Network. Vol. 279. Número 22. (pp. 1801-1807).
- Martín-Pliego, F. J., Ruiz-Maya, L. (2008). Fundamentos de inferencia estadística. Paraninfo.
- Mickey, R.M.; Durski, J.; Worden, J.K.: Danigelis, N.L. (1995). Breast cancer screening and associated factors for low-income African-American women. Preventive Medicine. Vol. 24. Issue 5. (pp.467-476).

- Organización Mundial de la Salud (2018). Cáncer. Disponible en <http://www.who.int/mediacentre/factsheets/fs297/es/>
- Perry, N.; Broeders, M.; DE Wolf, C.; Törnberg, S.; Holland, R. & Vonkarsa, L. (2008). European guidelines for quality assurance in breast cancer screening and diagnosis. *Annals of Oncology*. Vol.19. Issue 4. (pp.614–622).
- Red Española de Registros de Cáncer (2014). Estimaciones de la incidencia y la supervivencia del cáncer en España y situación en Europa. Disponible en <http://redecana.org/es/page.cfm?id=196&title=estimaciones-de-la-incidencia-y-la-supervivencia-del-cancer-en-espana-y-su-situacion-en-europa>
- Red Española de Registros de Cáncer (2016). Cancer incidence in Spain, 2015. Disponible en <http://redecana.org/es/page.cfm?id=196&title=estimaciones-de-la-incidencia-y-la-supervivencia-del-cancer-en-espana-y-su-situacion-en-europa>
- Sánchez, L. (2012). El amor lo cura todo: la importancia de las relaciones sociales en la salud. Disponible en <http://www.diariofemenino.com/salud/calidad-de-vida/articulos/amor-cura-todo-importancia-relaciones-sociales-salud/>
- SEER CANCER RESEARCH DATABASE (2017). National Cancer Institute of the United States.
- Serral, G.; Borrell, C.; Puigpinós I Riera, C (2018). Desigualdades socioeconómicas en el control mamográfico en mujeres españolas de 45 a 69 años de edad. Disponible en <https://www.sciencedirect.com/science/article/pii/S0213911117300390>
- Shinagawa, S.M. (2000). The excess burden of breast carcinoma in minority and medically underserved communities: application, research, and redressing institutional racism. *Cancer*. Vol. 88. Suppl. 5. (pp.1217-1223)
- Sociedad Española de Oncología Médica (2017a). Prevención. Disponible en <https://www.seom.org/es/informacion-sobre-el-cancer/prevencion-cancer>

- Sociedad Española de Oncología Médica (2017b): Cifras del cáncer en España en 2017. Disponible en [https://www.seom.org/seomcms/images/stories/recursos/Las cifras del cancer en Esp 2017.pdf](https://www.seom.org/seomcms/images/stories/recursos/Las_cifras_del_cancer_en_Esp_2017.pdf)
- Sociedad Española de Oncología Médica. Cifras del cáncer en España en 2018” Disponible en [https://seom.org/seomcms/images/stories/recursos/Las Cifras del cancer en Espana2018.pdf](https://seom.org/seomcms/images/stories/recursos/Las_Cifras_del_cancer_en_Espana2018.pdf)
- Stock, J.H. & Watson, M. (2012). Introducción a la Econometría. Editorial Pearson. (pp. 229-230).
- Universidad Nacional del Nordeste (2000). Incidencia de Cáncer de Mama en nuestra población. Estudio Preliminar. Disponible en http://www.unne.edu.ar/unnevieja/Web/cyt/cyt/2000/3_medicas/m_pdf/m_051.pdf
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*. (pp.817-838).
- Wingo, P.A.; Ries, L.A.; Giovino, G.A.; Miller, D.S.; Rosenberg, H.M. & Shopland, D.R. (1999). Annual report to the nation on the status of cancer, with a special section on lung cancer and tobacco smoking. *Journal of the National Cancer Institute*. Vol. 91. Issue 8. (pp.675–690).
- World Cancer Research Fund (2012). Link between lifestyle & cancer risk y Worldwide data. Disponible en <http://www.wcrf.org/>
- World Cancer Research Fund (2015). Breast cancer statistics. Disponible en <https://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics>.