

**Un estudio de la frecuencia anual de defunciones
acumuladas en cánceres específicos por edad
medido en el periodo
2013-2018**

Proyecto Área de Síntesis y Evaluación I (ASE I)

Universidad Iberoamericana

Alfredo Sandoval Rubalcava

Actuaría

Semestre otoño 2020

Lunes 30 de noviembre

Índice

1.-Planteamiento del problema:	3
2.-Objetivos:.....	4
3.- Justificación:.....	5
4.-Marco teórico:	5
5.-Metodología de la investigación	9
5.1.-Muestra	9
5.2.-Instrumentos de recopilación y manejo de datos:	11
5.3.-Herramientas a utilizar:.....	11
5.4.-Secuencia de pasos a seguir:	12
6.-Exploración de datos (resultados):	13
6.1.-Pronósticos y conclusión:	24
7.-Referencias	26
8.-Anexos	28

Un estudio de la frecuencia anual de defunciones acumuladas en cánceres específicos por edad medido en el periodo 2013-2018

1.-Planteamiento del problema:

- **Antecedentes:**

El cáncer es una enfermedad cuya frecuencia ha estado aumentando a lo largo de los años. En el año 2012 se registraron alrededor de 14 millones de nuevos casos y 8.2 millones de personas fallecidas (OMS, 2014). Para 2018, estas cifras ascendieron a 18.1 millones de casos nuevos y 9.6 millones de muertes (OMS, 2018).

A nivel mundial, en el año 2015, se atribuyeron a esta enfermedad alrededor de 8.8 millones de defunciones. Los 5 tipos de cáncer que causaron un mayor número de fallecimientos a nivel mundial son los siguientes:

1. Pulmonar (1,69 millones de defunciones)
2. Hepático (788 000 defunciones)
3. Colorrectal (774 000 defunciones)
4. Gástrico (754 000 defunciones)
5. Mamario (571 000 defunciones)

Ésta es de las principales causas de muerte en el mundo junto a la cardiopatía isquémica y el accidente cerebrovascular.

En esencia, esta enfermedad es una de las problemáticas principales que acechan a la humanidad y el propósito de este estudio es comprender su dinámica un poquito más.

A continuación, se explican antecedentes de los cánceres a estudiar en este trabajo.

Cáncer de hígado o hepático: La incidencia anual del cáncer hepático (el número de casos diagnosticados de forma anual) no se distribuye de la misma forma en todas las regiones del mundo. En países asiáticos y África subsahariana, éste llega a ser un tipo de tumor muy frecuente, sino el que más. Con base en los datos de Globocan, se han diagnosticado en 2018, aproximadamente 841,080 casos de cáncer hepático en el mundo, con 781,631 muertes causadas por este mal. Asimismo, este es el doble de frecuente en hombres que en mujeres. Ocupa el 9° lugar en incidencia de mujeres, 5° en hombres y 6° considerando ambos sexos (AECC,2019).

Cáncer de pulmón: El siglo pasado se consideraba al cáncer de pulmón como una enfermedad muy poco frecuente. En 1930 empezó a aumentar su frecuencia y hoy en día se considera el tumor maligno más frecuente en todo el mundo. En México, se encuentra entre las 3 principales causas de muerte por tumor maligno en adultos mayores de 35 años y es más frecuente en varones. Este cáncer está claramente

asociado con el consumo de tabaco ya que aproximadamente 80% de los casos se atribuyen a este hábito. Su incidencia es cercana a la mortalidad y se espera que crezca en los siguientes años (Moctezuma y Patiño, 2009).

Cáncer de mama: Este cáncer es la segunda causa de muertes por procesos tumorales en mujeres entre 1 y 21 años Según la CONAPO, Para el año 2020 se calcula que habrá 38.8 millones de casos, aproximadamente un tercio de la población del país (SCIELO, 2009).

A causa del crecimiento poblacional de mujeres entre 45 y 54 años, así como de 35 a 44 años, la tasa de mortalidad por cáncer de mama aumentó durante la última década, de 13.06% en 1990 a 14.49 durante el año 2000. Se puede apreciar una tendencia creciente de cáncer de mama en incidencia y mortalidad (SCIELO, 2009).

Cáncer cervicouterino: Este cáncer es uno de los padecimientos más frecuentes en la población femenina del mundo. Se calcula que anualmente se diagnostican 466 mil nuevos casos de este mismo y se producen alrededor de 300 mil muertes al año. Éste representa la segunda causa de muerte en países subdesarrollados (Medwave,2012).

Para el caso del país de México, este cáncer representa la segunda causa de muerte en mujeres por tumores malignos. Se puede relacionar este cáncer al factor de riesgo más común como el virus del papiloma humano (Medwave,2014).

Leucemia: En México hay aproximadamente cinco y seis mil casos de cáncer nuevos al año en menores de 18 años. De éstos, el 50% son debido a leucemias, afirma el jefe del servicio de oncología pediátrica, del Hospital Juárez de México (HJM), Gabriel Peñaloza González (SSA, 2019).

Este cáncer es el más común en la infancia y en la adolescencia. Se puede curar en muchos de los casos. El tiempo promedio que dura este tratamiento es de más de 2 años y medio y se somete a vigilancia médica el paciente durante 5 años. Si no hay reincidencia, se considera curado (SSA,2018).

- **Interrogante:** ¿Cómo se comporta la frecuencia acumulada de defunciones por edad medido anualmente durante el periodo 2013-2018 y qué se puede inferir de estos resultados; se pueden apreciar patrones o tendencias de estos con respecto a sus factores de riesgo?

2.-Objetivos:

- **Objetivo general:** Comprender cómo se comporta la frecuencia acumulada de defunciones por edad medido anualmente durante el periodo 2013-2018 e identificar patrones o tendencias en estas estos datos.
- **Objetivos específicos:**

- I. Identificar la distribución de defunciones por edad medida anualmente durante el periodo 2013-2018.
- II. Correlacionar estos datos y su distribución con las causas de estas enfermedades y sus factores de riesgo.
- III. Identificar patrones en los tipos de cáncer especificados en la investigación con los datos obtenidos.
- IV. Comparar los resultados obtenidos con los datos públicos para corroborar su veracidad.
- V. Proponer distintas vertientes de investigación con los resultados obtenidos, para poder entender el comportamiento de estos datos.

3.- Justificación:

Este tema me interesa ya que esta enfermedad ha estado presente en parientes cercanos a lo largo de mi vida y creo que poder comprender más a fondo cómo se comporta puede beneficiar a muchas personas, desde el sector asegurador hasta el sector salud.

4.-Marco teórico:

A continuación, se muestran definen los términos anatómicos de las enfermedades a tratar en el trabajo y los órganos especificados en los cánceres. Posteriormente se encuentran las definiciones matemáticas, así como las fórmulas, de los resultados estadísticos obtenidos en este trabajo para la parte inferencial.

Definiciones generales:

Pulmones: Estos son órganos vitales que permiten la respiración. Se absorbe O_2 y se expulsa CO_2 . Cada uno de los pulmones está formado por conductos que se hacen cada vez más estrechos: desde los bronquios hasta los alveolos, donde se produce el intercambio de estos gases. El oxígeno llega por las fosas nasales, faringe y laringe. Cuando llega a los pulmones hay un intercambio en la sangre para que se obtenga el O_2 para el cuerpo humano (MSD,2019).

Hígado: Este órgano está en la parte superior derecha del abdomen y es protegido por las costillas. Este órgano es vital y cumple una gran variedad de funciones.

- Produce bilis, sustancia que auxilia en la absorción de grasa por el organismo.
- Produce proteínas y coagulantes necesarios.
- Regula el azúcar (glucosa) en la sangre y almacena la adicional.
- Trabaja en conjunto con el estómago y el intestino para la digestión de los alimentos.
- Elimina y administra la producción de colesterol.

- Almacena grasas, minerales y vitaminas.
- Limpia la sangre, eliminando sustancias tóxicas y venenosas. También medicamentos recetados y de venta libre.
- Conformar una parte del sistema inmunológico, que ayuda a su cuerpo para combatir infecciones (Fairview, 2020).

Útero: Es un órgano muscular hueco que se ubica en la pelvis femenina, entre la vejiga y el recto. En éste, los ovarios producen los óvulos para ser fertilizados después de pasar por las trompas de Falopio, una vez que el óvulo se desprende del ovario al ser fertilizado y se adhiere a las paredes del útero. Su función principal es alimentar al feto en desarrollo previo al nacimiento (Medline Plus, 2020).

Mama o glándula mamaria: Es el órgano glandular ubicado en el pecho. Este consiste en tejido conjuntivo, tejido mamario y grasa. Estas glándulas producen la leche materna (INC, 2019).

Sangre: Es tejido vivo que tiene sólidos y líquidos. La parte sólida contiene glóbulos rojos, blancos y plaquetas. La otra, viene siendo el plasma que contiene agua, sales y proteínas (Medline Plus, 2020).

Los glóbulos blancos combaten las infecciones y son parte del sistema inmunológico del cuerpo. Los glóbulos rojos suministran oxígeno a los tejidos y órganos (Medline Plus, 2020).

Cáncer: “El término cáncer engloba un grupo numeroso de enfermedades que se caracterizan por el desarrollo de células anormales, que se dividen, crecen y se diseminan sin control en cualquier parte del cuerpo” (SEOM, 2019).

Cáncer de hígado o hepático: “El cáncer de hígado es consecuencia del desarrollo de células tumorales (es decir, que se dividen de forma anormal e incontrolada). Estas células tienen la capacidad de invadir los tejidos sanos de alrededor en el propio hígado y de crecer en órganos alejados e implantarse en ellos; es decir, de producir metástasis en otros órganos” (AECC,2019).

“Podemos distinguir entre el cáncer de hígado originado en el propio órgano o primario y el cáncer que asienta en el hígado pero que proviene de otros tumores denominados secundarios o cáncer de hígado metastásico” (AECC,2019).

Cáncer de pulmón: “El cáncer de pulmón generalmente se forma en las células que recubren los conductos de aire en los pulmones. Los dos tipos principales son el cáncer pulmonar de células pequeñas (normalmente presente en los fumadores de cigarrillos) y el cáncer pulmonar de células no pequeñas” (Radiology,2019).

Cáncer de mama: “El *cáncer de mama* es una enfermedad en la cual las células de la mama se multiplican sin control. Existen distintos tipos de cáncer de mama. El tipo de cáncer de mama depende de qué células de la mama se vuelven cancerosas” (CDC,2020).

“El cáncer de mama puede comenzar en distintas partes de la mama. Las mamas constan de tres partes principales: lobulillos, conductos y tejido conectivo. Los lobulillos son las glándulas que producen leche. Los conductos son los tubos que transportan la leche al pezón. El tejido conectivo (formado por tejido fibroso y adiposo) rodea y sostiene todas las partes de la mama. La mayoría de los cánceres de mama comienzan en los conductos o en los lobulillos” (CDC,2020).

Cáncer cervicouterino: “El cáncer de cuello uterino es un tipo de cáncer que se produce en las células del cuello uterino, la parte inferior del útero que se conecta a la vagina” (Mayo Clinic, 2019).

“Varias cepas del virus del papiloma humano (VPH), una infección de transmisión sexual, juegan un papel importante en la causa de la mayoría de los tipos de cáncer de cuello uterino” (Mayo Clinic, 2019).

Leucemia: “Enfermedad de los órganos productores de la sangre que se caracteriza por la proliferación excesiva de leucocitos o glóbulos blancos en la sangre y en la médula ósea” (Oxford, 2020).

Variable aleatoria: Es una función matemática obtenida de un experimento aleatorio. Esta frase se divide en dos

- Función matemática: Ecuación que asigna a valores una variable (variable dependiente) en función de otra u otras variables (variables independientes)
- Experimento aleatorio: Fenómeno en la vida real, de donde sus resultados provienen del azar. Esto quiere decir que, si se repite el experimento con las mismas condiciones, puede haber un resultado distinto.

Con estos dos conceptos en mente, podemos entender que una variable aleatoria es aquella función que intenta describir o describe (con un número) los resultados de un evento o eventos que provienen del azar (López, 2019).

Distribución normal: “Es un modelo matemático teórico capaz de aproximar satisfactoriamente el valor de una variable aleatoria a una situación real” (Rodo,2020).

“En otras palabras, la distribución normal adapta una variable aleatoria a una función que depende de la media y la desviación típica. Es decir, la función y la variable aleatoria tendrán la misma representación, pero con ligeras diferencias” (Rodo,2020).

Media aritmética:

$$Media(X) = \bar{x} = \frac{\sum_{i=1}^N X_i}{N}$$

siendo (X_1, X_2, \dots, X_N) el conjunto de observaciones

Imagen 1: Media aritmética (Requena, 2014)

Varianza:

$$S_X^2 = \frac{\sum_{i=1}^N (X_i - \bar{x})^2}{N - 1}$$

siendo (X_1, X_2, \dots, X_N) un conjunto de datos y \bar{x} la media

Imagen 2: Varianza (Requena, 2014)

Coeficiente de asimetría de Fisher:

$$CA_F = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N \cdot S_x^3}$$

siendo \bar{x} la media y S_x la desviación típica

Imagen 3: Coeficiente de asimetría de Fisher (Requena, 2014)

Curtosis:

$$Curtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{N \cdot S_x^4} - 3$$

siendo \bar{x} la media y S_x la desviación típica

Imagen 4: Curtosis (Requena, 2014)

Media marginal:

$$\bar{x}_j = \frac{\sum_{i=1}^l X_{ji}}{N}$$

Desviación estándar marginal:

$$\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^l (x_{ji} - \bar{x}_j)^2}{N}}$$

Covarianza:

$$\sigma_{x_j x_{j+1}} = \frac{\sum_{i=1}^l (x_{ji} x_{j+1i} - \bar{x}_j \bar{x}_{j+1})}{N}$$

Coeficiente de correlación de Pearson:

$$r = \frac{\sigma_{x_j x_{j+1}}}{\sigma_{x_j} \sigma_{x_{j+1}}}$$

5.-Metodología de la investigación

5.1.-Muestra

La muestra establecida en este trabajo es la frecuencia de muertes acumuladas de 5 tipos de cáncer distintos anualmente en el periodo 2013-2015. Es importante notar que los cánceres vienen con un código de enfermedad y los estudiados en este trabajo son tumores malignos. A continuación, se muestra el intervalo de cánceres a estudiar excluyendo aquellos dentro de este mismo que no son tumores malignos.

-Hígado (c220-c249)

- C220 Carcinoma de células hepáticas
- C221 Carcinoma de vías biliares intrahepáticas
- C222 Hepatoblastoma
- C223 Angiosarcoma del hígado
- C224 Otros sarcomas del hígado

- C227 Otros carcinomas especificados del hígado
- C229 "Tumor maligno del hígado, no especificado"
- C23X Tumor maligno de la vesícula biliar
- C240 Tumor maligno de las vías biliares extrahepáticas
- C241 Tumor maligno de la ampolla de Vater
- C249 "Tumor maligno de las vías biliares, parte no especificada"

-Pulmón (c340-c349)

- C340 Tumor maligno del bronquio principal
- C341 "Tumor maligno del lóbulo superior, bronquio o pulmón"
- C342 "Tumor maligno del lóbulo medio, bronquio o pulmón"
- C343 "Tumor maligno del lóbulo inferior, bronquio o pulmón"
- C349 "Tumor maligno de los bronquios o del pulmón, parte no especificada"

-Mama (c500-c509)

- C500 Tumor maligno del pezón y aréola mamaria
- C501 Tumor maligno de la porción central de la mama
- C502 Tumor maligno del cuadrante superior interno de la mama
- C503 Tumor maligno del cuadrante inferior interno de la mama
- C504 Tumor maligno del cuadrante superior externo de la mama
- C505 Tumor maligno del cuadrante inferior externo de la mama
- C506 Tumor maligno de la prolongación axilar de la mama
- C508 Lesión de sitios contiguos de la mama
- C509 "Tumor maligno de la mama, parte no especificada"

-Cervicouterino (c530-c55X)

- C530 Tumor maligno del endocérnix
- C531 Tumor maligno del exocérnix
- C538 Lesión de sitios contiguos del cuello del útero
- C539 "Tumor maligno del cuello del útero, sin otra especificación"
- C540 Tumor maligno del istmo uterino
- C541 Tumor maligno del endometrio
- C542 Tumor maligno del miometrio
- C543 Tumor maligno del fondo del útero
- C549 "Tumor maligno del cuerpo del útero, parte no especificada"
- C55X "Tumor maligno del útero, parte no especificada"

-Leucemia (c901-c959)

- C901 Leucemia de células plasmáticas
- C910 Leucemia linfoblástica aguda [LLA]

- C911 Leucemia linfocítica crónica de células tipo B
- C912 Leucemia linfocítica subaguda
- C913 Leucemia prolinfocítica de células tipo B
- C914 Leucemia de células vellosas
- C915 Leucemia /linfoma de células T adultas [HTLV-1-asociado]
- C916 Leucemia prolinfocítica de células tipo T
- C917 Otras leucemias linfoides
- C918 Leucemia tipo Burkitt de células B maduras
- C919 "Leucemia linfoide, sin otra especificación"
- C920 Leucemia mieloblástica aguda [LMA]
- C921 "Leucemia mieloide crónica [LMC], BCR/ABL-positivo"
- C922 "Leucemia mieloide crónica atípica, BCR/ABL-negativo"
- C924 Leucemia promielocítica aguda [LPA]
- C925 Leucemia mielomonocítica aguda
- C926 Leucemia mieloide aguda con anormalidad 11q23
- C927 Otras leucemias mieloides
- C928 Leucemia mieloide aguda con displasia multilinaje
- C929 "Leucemia mieloide, sin otra especificación"
- C930 Leucemia monocítica/monoblástica aguda
- C931 Leucemia mielomonocítica crónica
- C932 Leucemia monocítica subaguda
- C933 Leucemia mielomonocítica juvenil
- C937 Otras leucemias monocíticas
- C939 "Leucemia monocítica, sin otra especificación"
- C940 Leucemia eritroide aguda
- C942 Leucemia megacarioblástica aguda
- C943 Leucemia de mastocitos
- C947 Otras leucemias especificadas
- C950 "Leucemia aguda, células de tipo no especificado"
- C951 "Leucemia crónica, células de tipo no especificado"
- C952 "Leucemia subaguda, células de tipo no especificado"
- C957 Otras leucemias de células de tipo no especificado
- C959 "Leucemia, no especificada"

5.2.-Instrumentos de recopilación y manejo de datos:

Las herramientas utilizadas para filtrar los datos y manejarlos fueron: Excel, SQL y Python. Para las librerías de Python, se usaron Numpy, Pandas, Seaborn y Matplotlib Pyplot.

5.3.-Herramientas a utilizar:

Los datos para las defunciones vienen de los datos abiertos de la secretaria de salud. Se utilizó Excel y SQL para filtrar los fallecidos por enfermedad y Python para juntar las tablas. Asimismo, se utilizaron las herramientas de Mathematica y las librerías previamente mencionadas para poder generar las gráficas y las aproximaciones con los resultados obtenidos como parte del manejo de datos. Los códigos de Python se anexarán al final del trabajo en el anexo 1.

5.4.-Secuencia de pasos a seguir:

1.- Obtener los datos: Se buscaron los datos a partir de las bases de datos de defunciones de la secretaria de salud de México para poder obtener una muestra apropiada medida anualmente en el periodo 2013-2018.

2.-Filtrar los datos: La base de datos tenía defunciones de todo tipo y mi interés solo era los cánceres especificados previamente. Por lo que fue necesario filtrar por enfermedades y luego volver a filtrar por cáncer específico por año y edad. Todo esto fue en Excel y SQL con las herramientas que se nos enseñaron en clase acerca de esta plataforma.

3.-Juntar los datos: Una vez obtenidos los datos por año y cáncer en específico como tablas “.csv” de defunciones, se requería juntar estas tablas en una sola por cada cáncer. Para esto se utilizaron las librerías de Python previamente mencionadas para trabajar con los archivos “.csv” y juntar todas para que quedara una sola tabla por cada cáncer por edad, que la representaría en los 6 años.

4.-Obtener la suma acumulada de las defunciones: Se regresó a la herramienta de SQL para juntar la suma acumulada de cada cáncer por edad en el periodo establecido. Una vez que se tenían todas las tablas juntas, ya sólo se tenía que generar una sola con la suma cumulada por edad ya que el periodo estaba ya hecho.

5.-Generar las gráficas: Una vez completadas las tablas de las defunciones acumuladas del periodo deseado, se utilizó Python y las librerías mencionadas previamente para poder generar las gráficas de las defunciones. Este código se encuentra en el anexo 2.

6.-Obtener la función que aproxima la muestra obtenida: Al ver que se comportaban los datos como una función normal, se utilizó la media aritmética y la varianza, así como la suma acumulada de las defunciones para poder generar una función normal que aproxime los datos obtenidos de la base de datos de la secretaría de salud. Las imágenes del archivo están en el análisis de resultados de este trabajo. Asimismo, se obtuvo su curtosis y coeficiente de asimetría para comprender el comportamiento de las aproximaciones. Posteriormente se calculó la media marginal, desviación estándar marginal, covarianza y correlación de dos pares de cánceres (hígado-pulmón y mama-cervicouterino) para relacionar e inferir estos pares de variables y

comprender si había alguna relación matemática en estas como lo sugieren las gráficas previamente mencionadas.

7.-Inferir a partir de los datos obtenidos: ver si los datos obtenidos concuerdan con lo mostrado en los antecedentes y generar una respuesta a la interrogante para poder ver si se cumplió el objetivo general y los objetivos particulares.

6.-Exploración de datos (resultados):

esta exploración de datos comenzará con una vista general al panorama de las defunciones por edad acumulada en el periodo de todos lo cáncer específicos y posteriormente se analizarán uno por uno con sus respectivas inferencias a partir de lo enunciado en los objetivos, antecedentes y marco teórico.

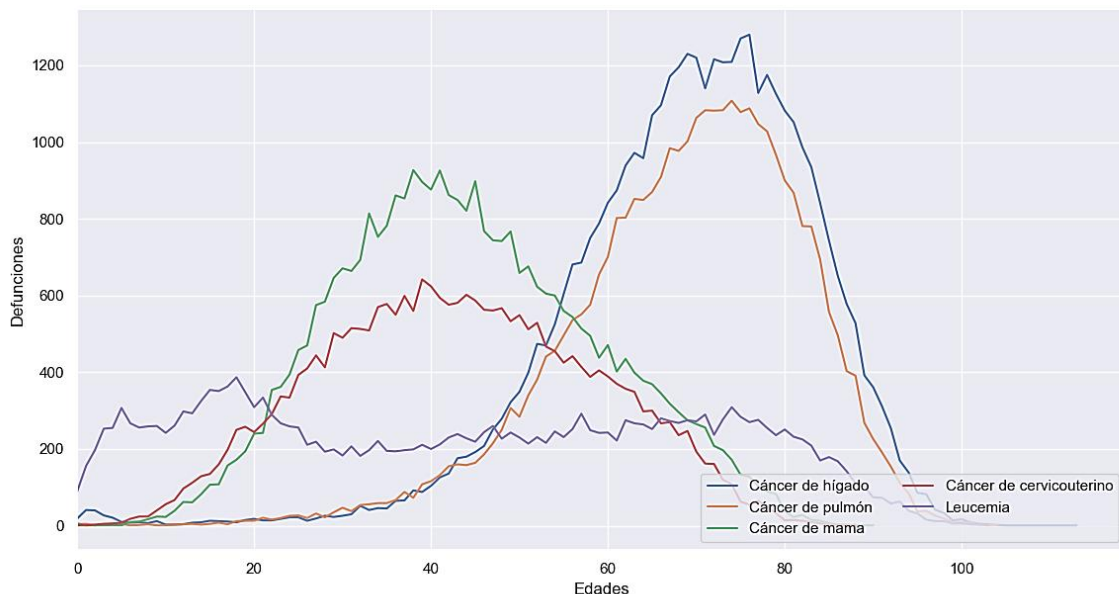


Figura 5: Defunciones acumuladas por cánceres específicos en el periodo 2013-2018

A grandes rasgos, se puede observar que cuatro de los cinco cánceres estudiados tienden a la forma de “campana” o función normal. Esto se discutirá más a fondo en cada cáncer en específico.

Como se menciona en los antecedentes, los cánceres predominantes en hombres; cáncer de pulmón e hígado, tienden a tener la cresta de sus respectivas campanas por el mismo rango de edades. Lo mismo ocurre con los cánceres predominantes en mujeres, el cervicouterino y de mama. Esto se puede ver con la correlación de Pearson obtenida para ambos pares de cánceres.

Correlación de Pearson para Cáncer de Hígado y de Pulmón: 0.9798

Error absoluto $|1-0.9798|= 0.0202$

Correlación de Pearson para Cáncer Cervicouterino y de Mama: 0.9658

Error absoluto $|1-0.9658|= 0.0342$

Podemos ver que ambas muestran una correlación lineal positiva fuerte ya que se acercan mucho al uno con errores absolutos muy cercanos al cero. En esta parte se puede concluir que, si se habla de una predominancia en casos de hombres y de mujeres para ciertos tipos de cáncer, es porque sí existe una correlación clara entre ambos pares de variables.

Cáncer de Hígado:

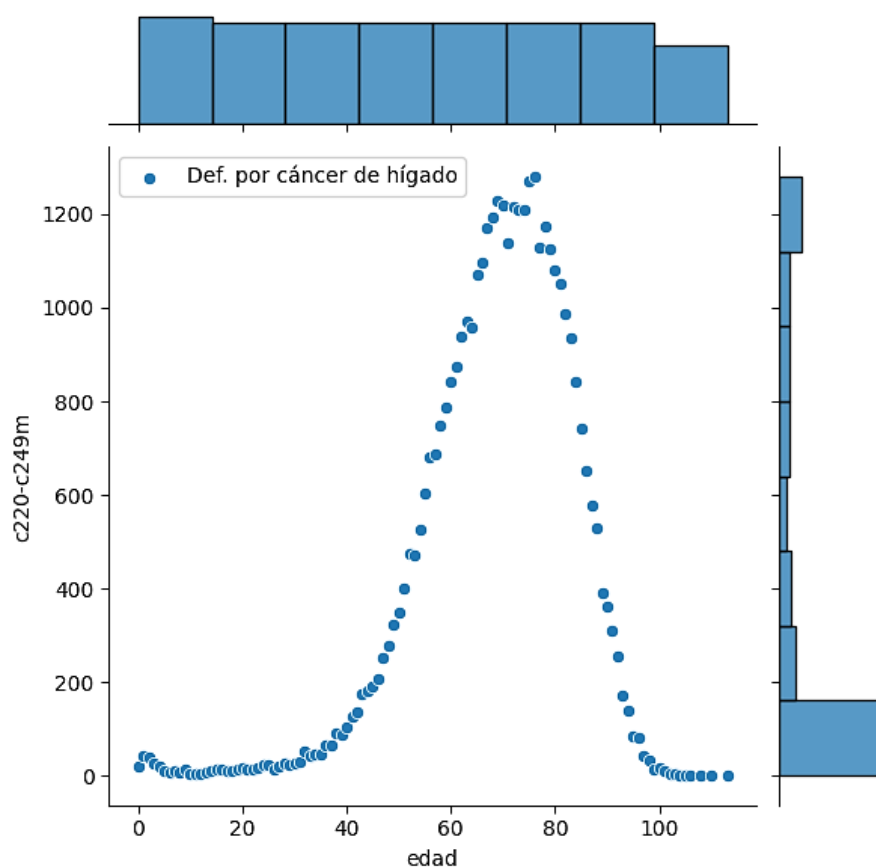


Figura 6: Defunciones acumuladas por cáncer de hígado en el periodo 2013-2018

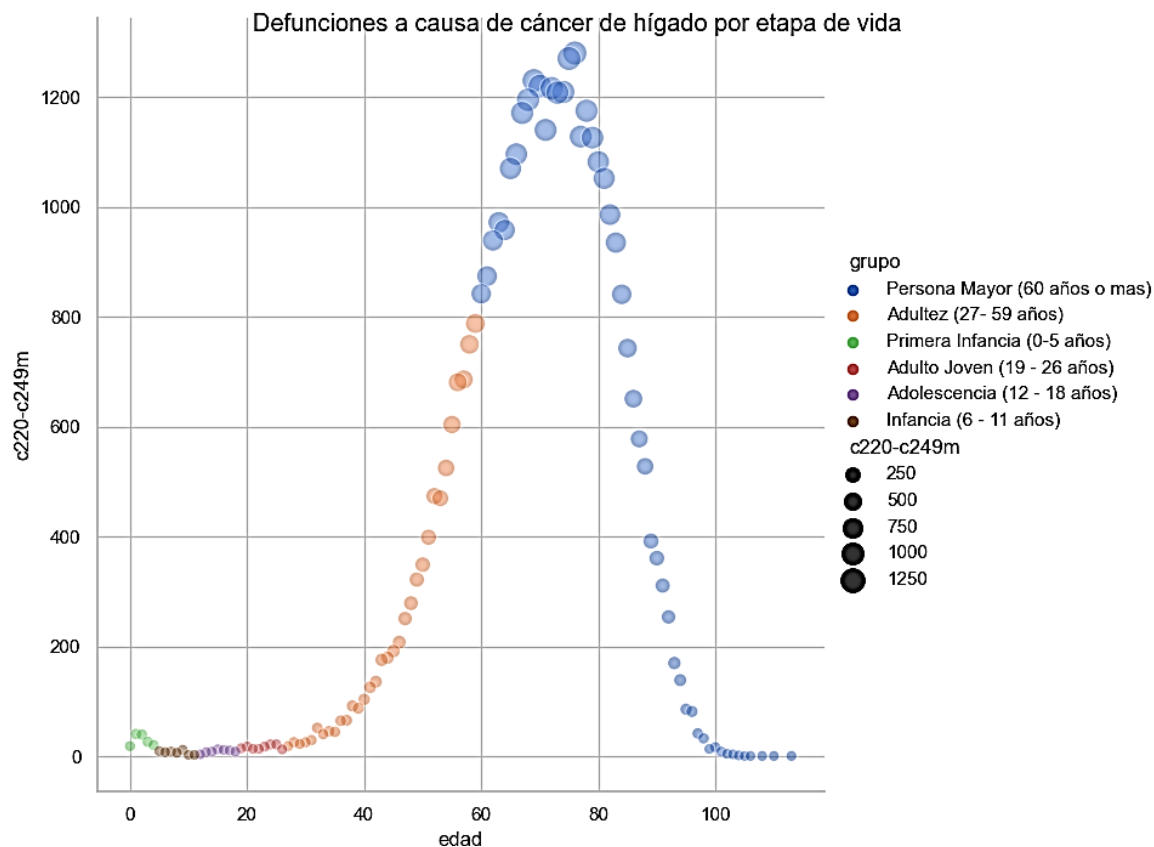


Figura 7: Defunciones acumuladas por cáncer de hígado por etapa de vida en el periodo 2013-2018

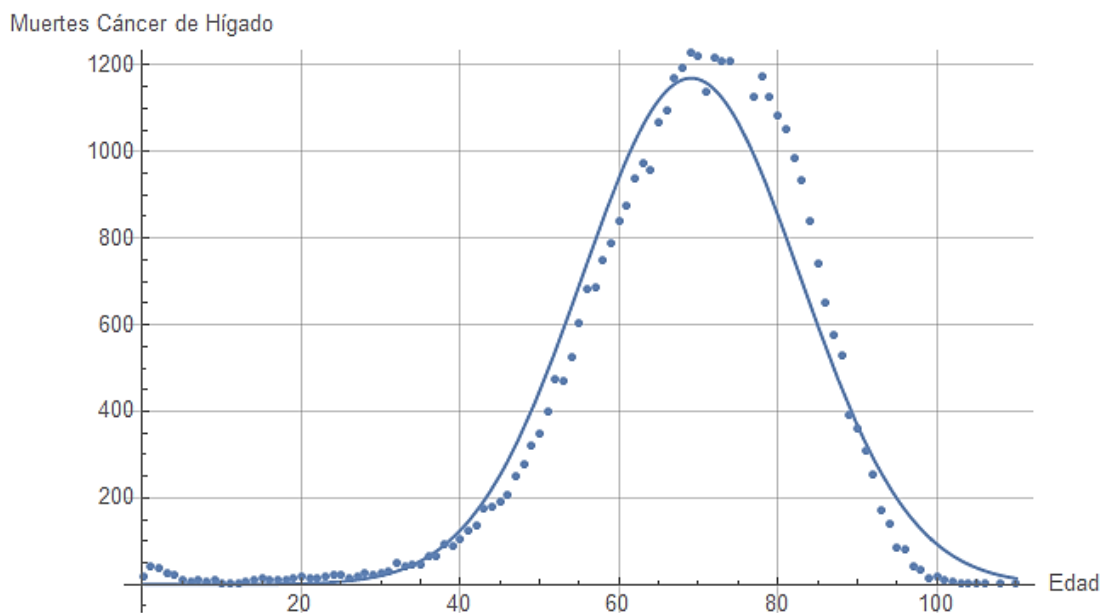


Figura 8: Aproximación de defunciones acumuladas por cáncer de hígado en el periodo 2013-2018

Se puede apreciar en la figura 6 cómo la acumulación de defunciones en la vertical es relativamente constante, mientras que en la horizontal tenemos un salto en los niveles bajos de defunciones, es decir, las colas tienen una cantidad de muertos mayor cuando se suman en el eje de las abscisas.

En la figura 7, se pueden apreciar los fallecimientos por etapa de vida. La mayor cantidad de difuntos ocurre en la etapa de la adultez y personas mayores.

Para la aproximación realizada en matemática, se puede apreciar una normal leptocúrtica ajustada con un coeficiente de asimetría negativo y los siguientes datos de estadística descriptiva.

Media	69.0852
Varianza	189.342
Desviación estándar	13.7601
Coeficiente de asimetría	-0.972396
Curtosis	2.39587

Con esta inferencia, se puede comentar que, si pensamos en los factores de riesgo de esta enfermedad, como lo es el alcoholismo (que provoca la predisposición al cáncer de hígado), no es de sorprenderse que, en México, las defunciones por esta enfermedad sean más que las de cáncer de pulmón.

Cáncer de Pulmón:

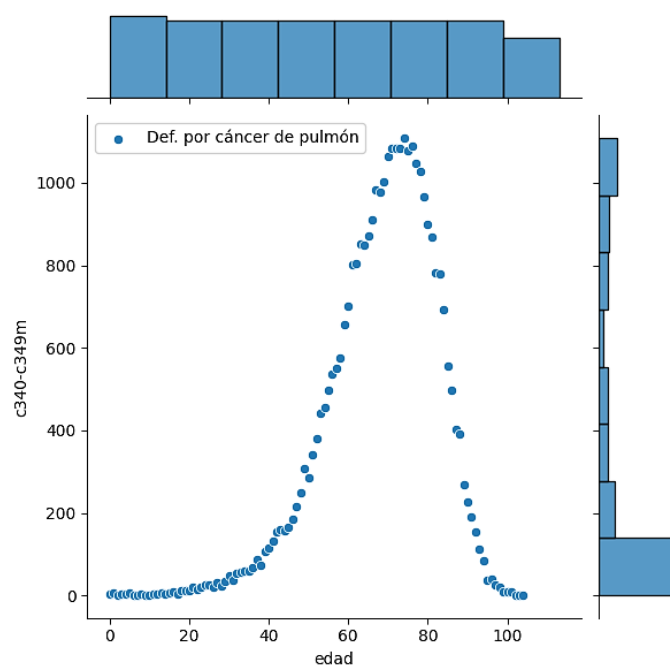


Figura 9: Defunciones acumuladas por cáncer de pulmón en el periodo 2013-2018

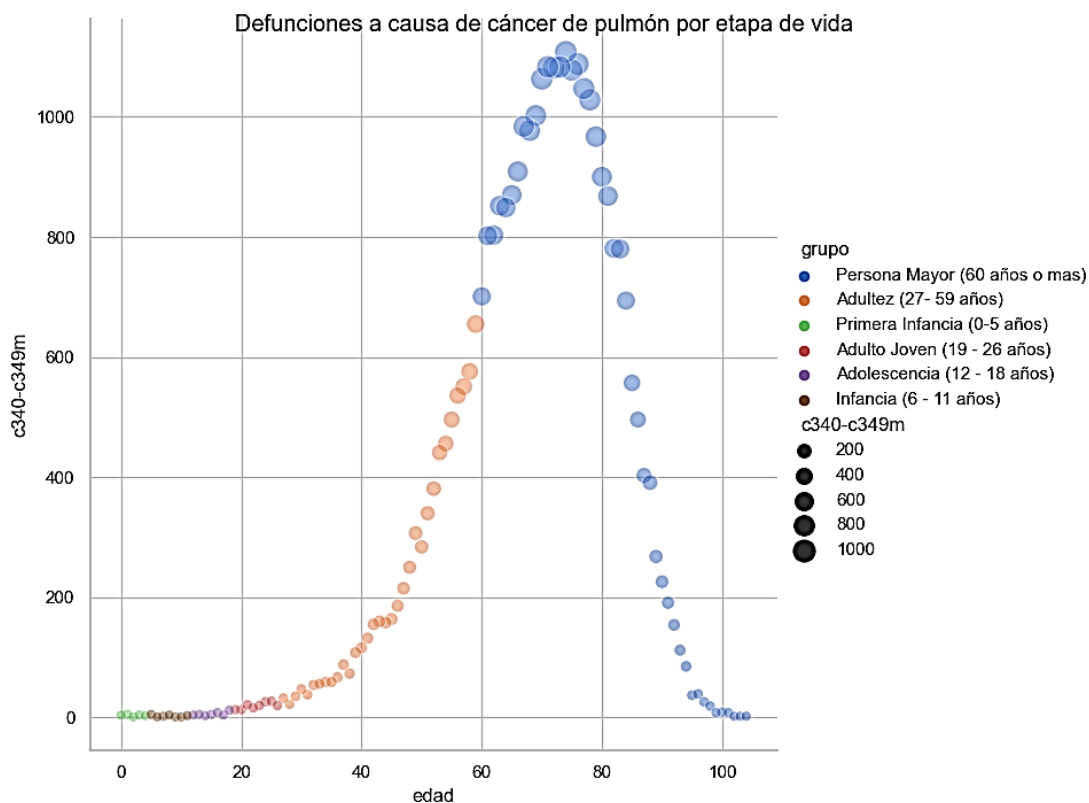


Figura 10: Defunciones acumuladas por cáncer de pulmón por etapa de vida en el periodo 2013-2018

Muertes Cáncer de Pulmón

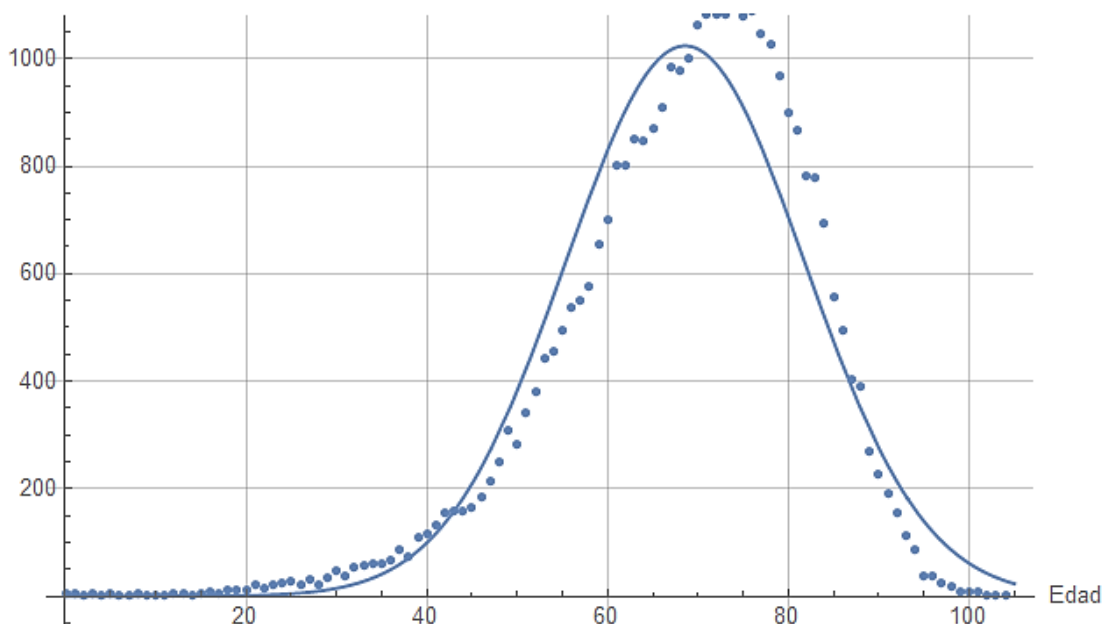


Figura 11: Aproximación de defunciones acumuladas por cáncer de pulmón por en el periodo 2013-2018

En la figura 9 podemos notar una forma similar al cáncer de hígado en el eje de la suma por el eje de las ordenadas y las abscisas. Un comportamiento constante por la vertical y un pico en las colas por la horizontal.

Al ubicarnos en la figura 10, comprendemos cómo, nuevamente, hay una mayor acumulación de fallecimientos en las etapas de adultez y personas mayores.

Para la aproximación realizada en Matemática, se puede apreciar una normal leptocúrtica ajustada con un coeficiente de asimetría negativo y los siguientes datos de estadística descriptiva.

Media	68.5746
Varianza	175.066
Desviación estándar	13.2313
Coeficiente de asimetría	-0.759834
Curtosis	1.01629

Con esto se pudo inferir que, aunque el cáncer de pulmón sea el más frecuente a nivel mundial, éste se encuentra por debajo del cáncer de Hígado. Cabe mencionar que a partir de los 27 años se vuelve muy frecuente este cáncer y aproximadamente el 80% de los casos se atribuyen al factor de riesgo del consumo del tabaco. En conclusión, fumar **SÍ** mata.

Cáncer de mama:

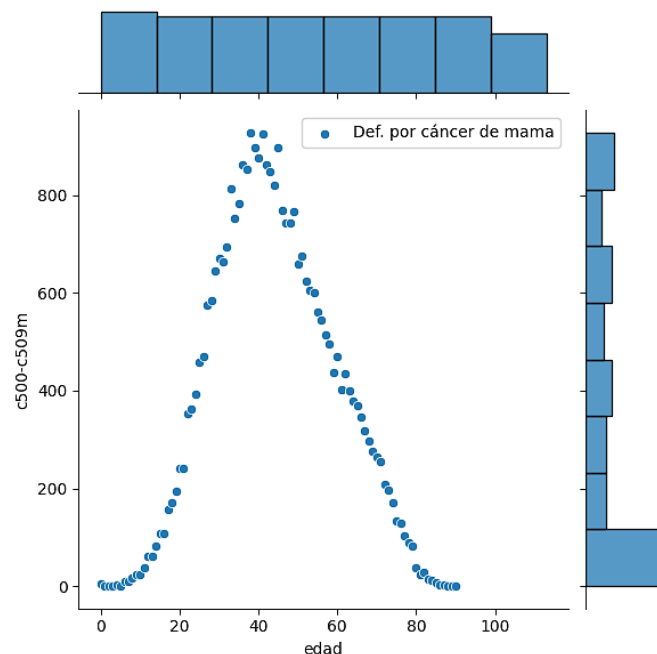


Figura 12: Defunciones acumuladas por cáncer de mama en el periodo 2013-2018

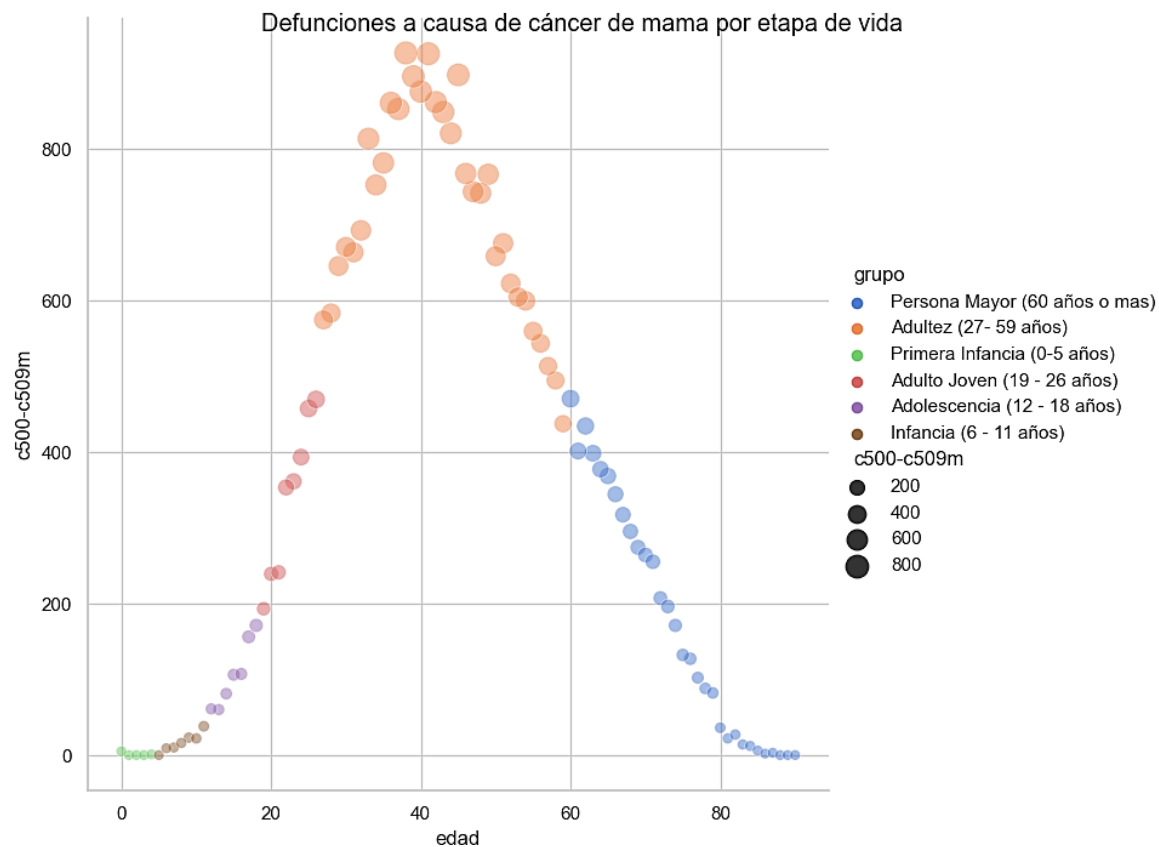


Figura 13: Defunciones acumuladas por cáncer de mama por etapa de vida en el periodo 2013-2018

Muertes Cáncer de mama

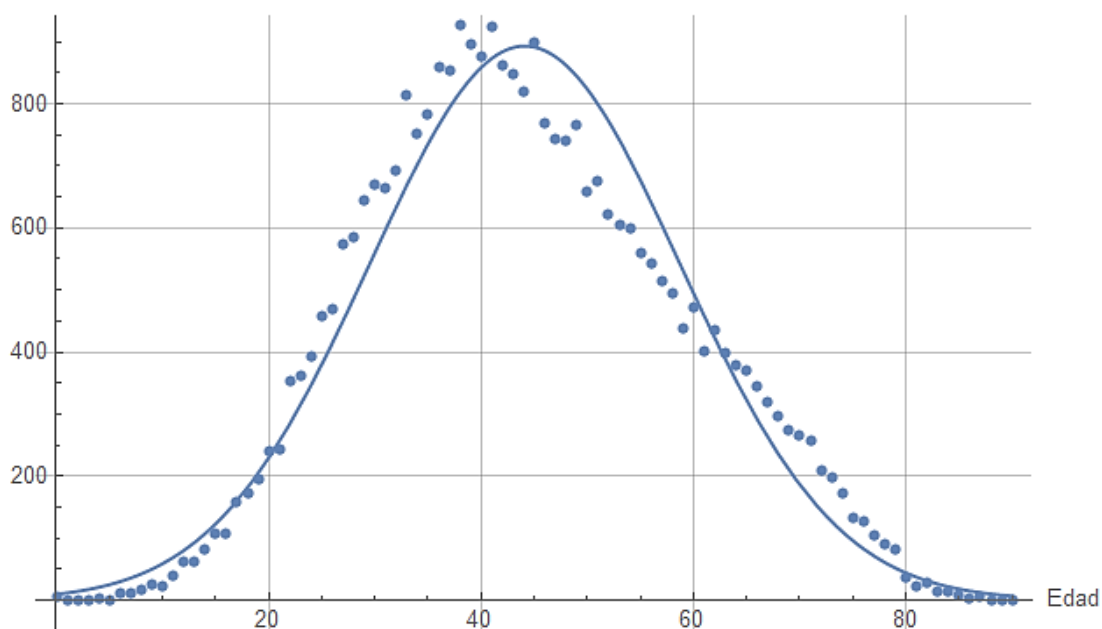


Figura 14: Aproximación de defunciones acumuladas por cáncer de pulmón en el periodo 2013-2018

Al ver la figura 11, se aprecia el comportamiento habitual hasta el momento, en el cual el eje de la suma por el eje de las ordenadas y las abscisas tienen las tendencias de uniformidad en las verticales y el pico en horizontales.

Viendo la figura 12, se puede apreciar como el cáncer de mama afecta mayormente a mujeres en la etapa adulta, pero comienza a crecer desde la etapa de adulto joven.

Para la aproximación realizada en Matemática, se puede apreciar una normal platicúrtica ajustada con un coeficiente de asimetría positivo y los siguientes datos de estadística descriptiva.

Media	44.1075
Varianza	213.672
Desviación estándar	14.6173
Coeficiente de asimetría	0.22224
Curtosis	-0.459913

Se infiere a partir de estos datos que el cáncer de mama no presenta tantas defunciones en el rango de edades de 1 a 21 comparado con el cáncer cervicouterino. Sin embargo, a partir de este rango de edad, el crecimiento en defunciones se acelera de forma considerable, dejando corto al cáncer previamente mencionado. Los factores de riesgo para esta enfermedad son la obesidad, diabetes, predisposición hereditaria o factores hormonales.

Cáncer cervicouterino:

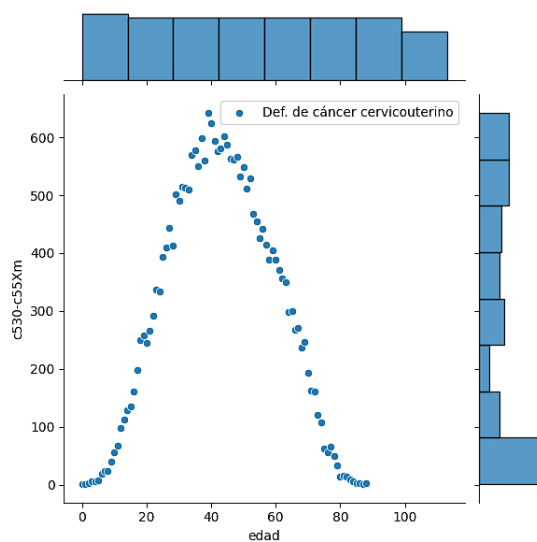


Figura 15: Defunciones acumuladas por cáncer cervicouterino en el periodo 2013-2018

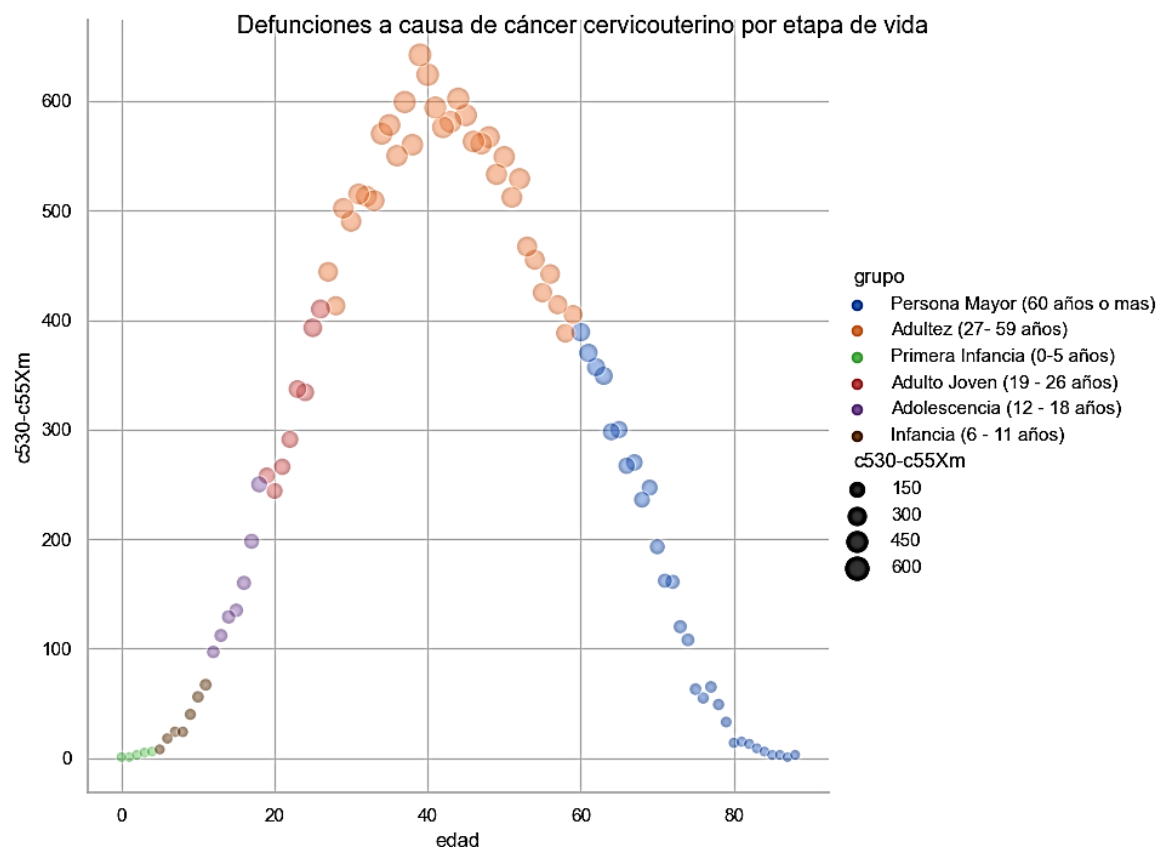


Figura 16: Defunciones acumuladas por cáncer cervicouterino por etapa de vida en el periodo 2013-2018

Muertes Cáncer Cervicouterino

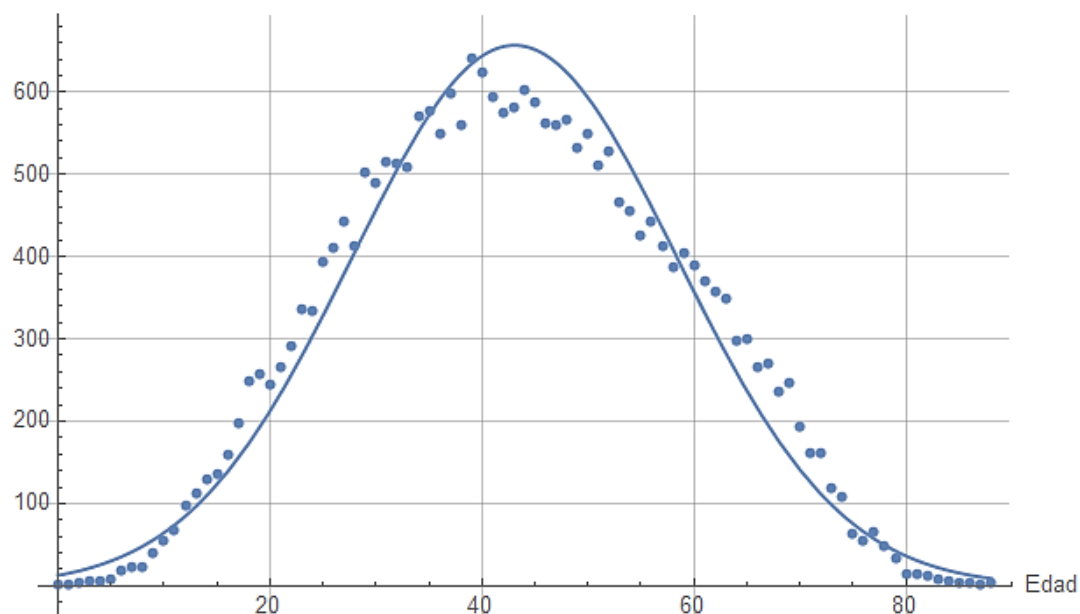


Figura 17: Aproximación de defunciones acumuladas por cáncer cervicouterino en el periodo 2013-2018

Nuevamente, en la figura 15, se aprecia la tendencia en ambas sumas de los ejes. En la vertical, hay un comportamiento uniforme, y en la horizontal se encuentra la mayor cantidad acumulada de los datos en las colas.

En la figura 16, hay una clara relación con el cáncer de mama ya que nuevamente hay una tendencia de mayor cantidad de muertes en la etapa de adulto joven y ya en la adultez se encuentra la mayor acumulación.

Para la aproximación realizada en Matemática, se puede apreciar una normal platicúrtica ajustada con un coeficiente de asimetría positivo y los siguientes datos de estadística descriptiva.

Media	43.0734
Varianza	235.766
Desviación estándar	15.3547
Coeficiente de asimetría	0.0641071
Curtosis	-0.621391

Aquí podemos inferir que esta es la segunda causa de muertes por tumores malignos en contraste con el cáncer de mama. Sin embargo, se puede ver la relación en ambas enfermedades por las edades en las cuales hay más defunciones. El factor de riesgo predominante en esta enfermedad es el virus del papiloma humano.

Leucemia:

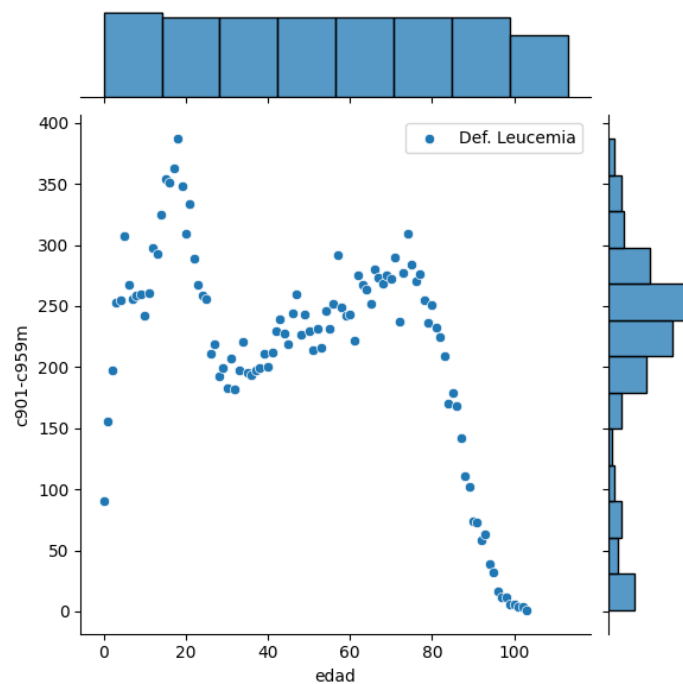


Figura 18: Defunciones acumuladas por leucemia en el periodo 2013-2018

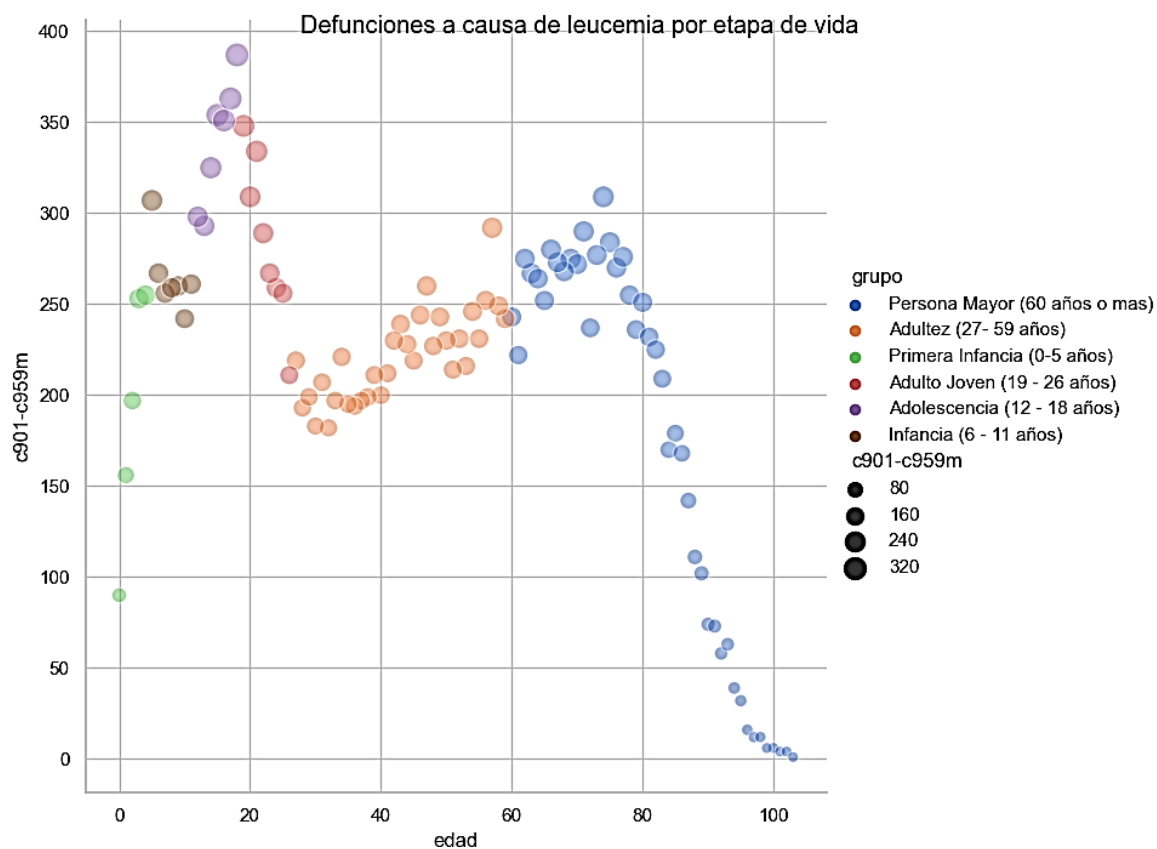


Figura 19: Defunciones acumuladas por leucemia por etapa de vida en el periodo 2013-2018

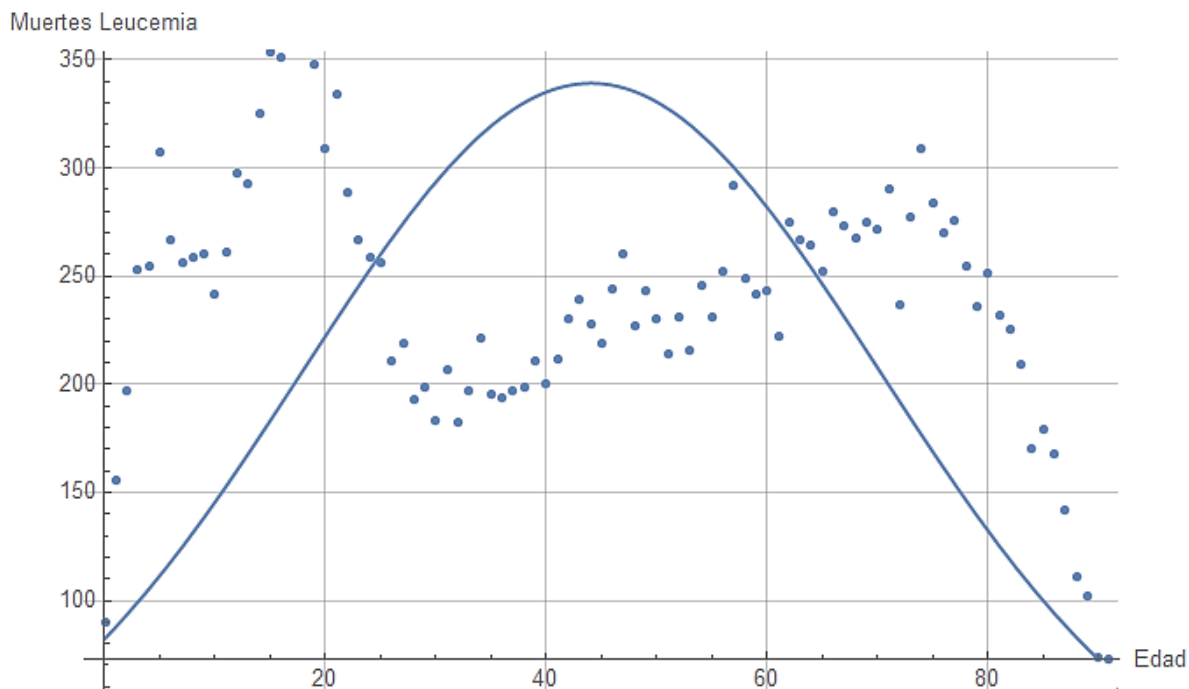


Figura 20: Aproximación de defunciones acumuladas por leucemia en el periodo 2013-2018

La figura 18 cambia por completo con respecto a las demás enfermedades. Podemos ver que en el eje de las ordenadas hay una acumulación constante. Sin embargo, para el eje de la horizontal, se pueden apreciar varios picos formando algo que pareciera ser una cuasinormal.

En la figura 19, hay se puede apreciar una clara tendencia de defunciones en las etapas de vida de primera infancia a la adolescencia. Como se decía previamente, este cáncer tiene más mortalidad en las edades tempranas que las demás.

Para la aproximación realizada en matemática, se puede apreciar una normal platicúrtica ajustada con un coeficiente de asimetría positivo y los siguientes datos de estadística descriptiva. Sin embargo, es muy importante entender que este modelo no se acopla a la función bimodal que se aprecia a los datos obtenidos. Esto puede ser sujeto a futura investigación en este proyecto para buscar la aproximación a esta bimodal.

Media	44.056
Varianza	685.557
Desviación estándar	26.1831
Coeficiente de asimetría	0.0545332
Curtosis	-1.25129

La inferencia para esta enfermedad es la apreciación de la mortalidad infantil. Aunque el factor de riesgo para esta enfermedad es la herencia o la mala nutrición, se sabe poco en comparación con los otros cánceres de cual es la causa exacta de esta misma. Del mismo modo, podemos ver que no hay tantas defunciones en proporción a los demás cánceres, pero la que hay es a muy temprana edad de la vida de las personas.

6.1.-Pronósticos y conclusión:

Esta investigación tiene muchas ramas para futura investigación, por ejemplo, relacionar la diabetes y la obesidad (temas vistos en esta carrera por parte de la disciplina de la demografía), con el cáncer de mama ya que es uno de sus factores de riesgo. Correlacionar más a fondo los factores de riesgo de estas enfermedades con su letalidad y poder generar difusión y conocimiento para informar a la población que, por ejemplo, el alcohol y el cigarro en exceso llegan a ser un factor de riesgo en la población mexicana. Estas estadísticas no se deben tomar a la ligera; sino todo lo contrario y ver esto como una advertencia ante los años por venir.

Ver como estas estadísticas se comparan con las de otros países también puede ser una vía de investigación ya que puede haber sesgos en la información de los que no somos conscientes y comparar la forma de estudiar los datos es algo indispensable en el campo de la investigación estadística y la ciencia de datos.

Propongo que esta investigación se incline primero a los factores de riesgo de estas enfermedades antes de ver el panorama internacional de estas mismas. Comprender la casa de uno es de vital importancia antes de poder conocer el mundo. Preveo que a futuro esto podrá fungir como glosario histórico ante el incremento de letalidad de estas enfermedades que habrá en el futuro no tan lejano, ya sea por el crecimiento poblacional, cambio climático o cualquier factor de riesgo que tenga que afrontar la humanidad. Este documento servirá como testimonio previo a lo que está por venir ya que “Los muertos no mienten”.

7.-Referencias

- Terrase, V. (Ed.). (2018, September 12). Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018. Retrieved November 20, 2020, from https://www.iarc.fr/wp-content/uploads/2018/09/pr263_E.pdf
- G. (Ed.). (2019, June 13). El cáncer en el mundo y México. Retrieved November 20, 2020, from <https://www.infocancer.org.mx/?c=conocer-el-cancer>
- OMS, (2018, May 24). Las 10 principales causas de defunción. Retrieved November 20, 2020, from <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- OMS. (2014, December 05). OMS | Datos y cifras sobre el cáncer. Retrieved November 20, 2020, from <https://www.who.int/cancer/about/facts/es/>
- SOEM. (2019, December 16). ¿Qué es el cáncer y cómo se desarrolla? Retrieved November 20, 2020, from <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla>
- Oxford, L. (2020). Leucemia: Definición de Leucemia por Oxford Dictionaries en Lexico.com también significado de Leucemia. Retrieved November 20, 2020, from <https://www.lexico.com/es/definicion/leucemia>
- Radiological Society of North America (RSNA) and American College of Radiology (ACR). (2019, July 17). Cáncer pulmonar. Retrieved November 20, 2020, from <https://www.radiologyinfo.org/sp/info.cfm?pg=lung-cancer>
- Mayo Clinic. (2019, December 14). Cáncer de cuello uterino. Retrieved November 20, 2020, from <https://www.mayoclinic.org/es-es/diseases-conditions/cervical-cancer/symptoms-causes/syc-20352501>
- AECC. (2019). Cáncer de hígado: ¿Qué es? Retrieved November 20, 2020, from <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-higado>
- Moctezuma, C. R., & Patiño, M. (2009). Cáncer de pulmón. Retrieved November 20, 2020, from <https://www.medigraphic.com/pdfs/anaradmex/arm-2009/arm091e.pdf>
- Fernández, M. E., & Martini, E. V. (2009). El cáncer de mama en México: Evolución, panorama actual y retos de la sociedad civil. Retrieved November 20, 2020, from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342009000800025

- Moran, R., & Quintero, M. (2012, August 01). Antecedentes sobre el cáncer cervicouterino en México e importancia de la educación sexual en la prevención temprana en jóvenes y población rural. Retrieved November 20, 2020, from <https://www.medwave.cl/link.cgi/Medwave/Revisiones/Análisis/5453>
- Salud, S. (2018). 055. La leucemia representa el 50 por ciento de los casos de cáncer infantil. Retrieved November 20, 2020, from <https://www.gob.mx/salud/prensa/055-la-leucemia-representa-el-50-por-ciento-de-los-casos-de-cancer-infantil>
- Merck Sharp & Dohme de España, S. (2019). MSD Salud. Retrieved November 25, 2020, from <https://www.msdsalud.es/cuidar-en/cancer-pulmon/quieres-saber-mas/son-sirven-pulmones.html>
- The StayWell Company, F. (2020). Funcionamiento del hígado. Retrieved November 25, 2020, from https://www.fairview.org/sitecore/content/Fairview/Home/Patient-Education/Articles/Spanish/f/u/n/c/i/Funcionamiento_del_h%C3%ADgado_40220_espa%C3%B1ol
- Martínez, L. (2020, August 6). Útero: MedlinePlus enciclopedia médica ilustración. Retrieved November 25, 2020, from https://medlineplus.gov/spanish/ency/esp_imagepages/19263.htm
- NIH. (2018, April 27). Estadísticas del cáncer. Retrieved November 25, 2020, from <https://www.cancer.gov/espanol/cancer/naturaleza/estadisticas>
- NIH. (2020). Sangre: MedlinePlus en español. Retrieved November 25, 2020, from <https://medlineplus.gov/spanish/blood.html>
- Londoño, N., López, J., Martínez, E., Henry, Luz, Morales, F., . . . Carlos. (2020, October 10). Distribución normal. Retrieved November 25, 2020, from <https://economipedia.com/definiciones/distribucion-normal.html>
- Josefina, A., & López, J. (2019, January 28). Variable aleatoria - Definición, qué es y concepto. Retrieved November 25, 2020, from <https://economipedia.com/definiciones/variable-aleatoria.html>
- Serra, B., Dice, R., Rosario del Socorro Orozco Palacios dice:, Dice:, F., Febrero, R., Dice:, R., N. (2020, October 26). Media (promedio o media aritmética). Retrieved November 25, 2020, from <https://www.universoformulas.com/estadistica/descriptiva/media/>

Barranco, A. (2018, May 4). Defunciones - datos abiertos defunciones, gobierno de México.
 Retrieved November 20, 2020, from
<https://datos.gob.mx/busca/dataset/defunciones>

8.-Anexos

Anexo 1. Python “juntar tablas”

```
import
pandas
as pd

m13 = pd.read_csv('c900-c959m13.csv')
m14 = pd.read_csv('c900-c959m14.csv')
m15 = pd.read_csv('c900-c959m15.csv')
m16 = pd.read_csv('c900-c959m16.csv')
m17 = pd.read_csv('c900-c959m17.csv')
m18 = pd.read_csv('c900-c959m18.csv')

JuntosAB= m13.append(m14)
dfJuntosAB=pd.DataFrame(JuntosAB)
#print(dfJuntosAB)

JuntosABC= dfJuntosAB.append(m15)
dfJuntosABC=pd.DataFrame(JuntosABC)
#print(dfJuntosABC)

JuntosABCD= dfJuntosABC.append(m16)
dfJuntosABCD=pd.DataFrame(JuntosABCD)
#print(dfJuntosABCD)

JuntosABCDE= dfJuntosABCD.append(m17)
dfJuntosABCDE=pd.DataFrame(JuntosABCDE)
#print(dfJuntosABCDE)

JuntosABCDEF= dfJuntosABCDE.append(m18)
dfJuntosABCDEF=pd.DataFrame(JuntosABCDEF)
print(dfJuntosABCDEF)

dfJuntosABCDEF.to_csv (r'C:\Users\alqui\Desktop\Proyectos Finales\Ciencia de
datos\Tablas6añosMcxxx\c900-c959m13-18SinAcumular.csv', index = False,
header=True)
```

Anexo 2 (Inferencia_c220-c959m13-18ConAcumular)

```

import
pandas
as pd

import seaborn as sns
import matplotlib.pyplot as plt
from numpy import random

# Se lee el CSV, cambias a tu csv este archivo
c220_c959m13_18ConAcumular = pd.read_csv('c220-c959m13-
18ConAcumularGrupo.csv')
# Se imprime la cabeza del archivo para ver que se haya leído bien el CSV
#cabeza = c220_c959m13_18ConAcumular.head()
#print(cabeza)

# Se genera la estadística descriptiva de las edades y los cánceres columna
por columna y al final todas juntas
dfCancer = pd.DataFrame(c220_c959m13_18ConAcumular)
datosTotalesHigado      = dfCancer['c220-c249m'].describe()
datosTotalesPulmon      = dfCancer['c340-c349m'].describe()
datosTotalesMama        = dfCancer['c500-c509m'].describe()
datosTotalesCervicouterino = dfCancer['c530-c55Xm'].describe()
datosTotalesLeucemia     = dfCancer['c901-c959m'].describe()
#print(datosTotalesHigado)
#print(datosTotalesPulmon)
#print(datosTotalesMama)
#print(datosTotalesCervicouterino)
#print(datosTotalesLeucemia)

#normc220_c249m= random.normal(loc=366.927273,scale=442.252153,size=(110))
#print(normc220_c249m)

#sns.distplot(normc220_c249m,bins=20)
#sns.lineplot(x='edad', y=normc220_c249m, data=c220_c959m13_18ConAcumular)
#plt.show()

#sns.distplot(c220_c959m13_18ConAcumular['c220-c249m'],bins=20)

# Para graficar en ambos ejes del plano dando dos columnas numéricas
distintas

```

```

#sns.jointplot(x='edad', y='c220-c249m',
data=c220_c959m13_18ConAcumular,label="Def. por cáncer de hígado")
#sns.jointplot(x='edad', y='c340-c349m',
data=c220_c959m13_18ConAcumular,label="Def. por cáncer de pulmón")
#sns.jointplot(x='edad', y='c500-c509m',
data=c220_c959m13_18ConAcumular,label="Def. por cáncer de mama")
#sns.jointplot(x='edad', y='c530-c55Xm',
data=c220_c959m13_18ConAcumular,label="Def. de cáncer cervicouterino")
#sns.jointplot(x='edad', y='c901-c959m',
data=c220_c959m13_18ConAcumular,label="Def. Leucemia")
#plt.show()

#sns.set_theme( style='darkgrid')
#f, ax=plt.subplots(figsize=(150,100))
#defunciones= c220_c959m13_18ConAcumular.sort_values("c220-
c249m",ascending=False)
#print(defunciones)
##NOTA: EL PRIMER SUBSET DEBE SER MAYOR AL SEGUNDO, EN caso de que no, se
comerá la segunda gráfica, la primera
#sns.lineplot(x="edad",y="c220-c249m",data=defunciones,label="Cáncer de
hígado")
#Se crea el segundo subset a guardar en el subplot
#sns.lineplot(x="edad",y="c340-c349m",data=defunciones,label="Cáncer de
pulmón")
#Se crea el tercer subset a guardar en el subplot
#sns.lineplot(x="edad",y="c500-c509m",data=defunciones,label="Cáncer de
mama")
#Se crea el cuarto subset a guardar en el subplot
#sns.lineplot(x="edad",y="c530-c55Xm",data=defunciones,label="Cáncer de
cervicouterino")
#Se crea el quinto subset a guardar en el subplot
#sns.lineplot(x="edad",y="c901-c959m",data=defunciones,label="Leucemia")
#ax.legend(ncol=2, loc="lower right",frameon=True)
#ax.set(xlim=(0),ylabel="Defunciones",xlabel="Edades")
#sns.despine(top=True,right=True)
#plt.show()

#sns.set_theme(style="ticks")

```

```

#sns.jointplot(x="c340-c349m", y="c220-c249m", kind="hex",
color="#4CB391",data=defunciones)
#plt.show()

#El de la galaxia que combina 3 gráficas en 1, la X y la Y están igualadas a
las columnas deseadas del data set
#f, ax = plt.subplots(figsize=(6, 6))
#x=c220_c959m13_18ConAcumular['edad']
#y=c220_c959m13_18ConAcumular['c220-c249m']
#sns.scatterplot(x=x, y=y, s=5, color=".15",data=defunciones)
#sns.histplot(x=x, y=y, bins=50, pthresh=.1, cmap="mako",data=defunciones)
#sns.kdeplot(x=x, y=y, levels=5, color="w", linewidths=1,date=defunciones)
#plt.show()

#El parámetro hue de este es para la clasificación ej. mujer/hombre
#sns.set_theme( style='whitegrid')
#a=sns.relplot(x="edad", y="c340-c349m", size="c340-c349m",
#             sizes=(40,200), alpha=.5, palette="muted",hue='grupo',
#             height=7, data=defunciones)
#a.fig.suptitle("Defunciones a causa de cáncer de pulmón por etapa de vida")
#plt.show()

#sns.set_theme( style='whitegrid')
#b=sns.relplot(x="edad", y="c220-c249m", size="c220-c249m",
#             sizes=(40,200), alpha=.5, palette="muted",hue='grupo',
#             height=7, data=defunciones)
#b.fig.suptitle("Defunciones a causa de cáncer de hígado por etapa de vida")
#plt.show()

#sns.set_theme( style='whitegrid')
#c=sns.relplot(x="edad", y="c500-c509m", size="c500-c509m",
#             sizes=(40,200), alpha=.5, palette="muted",hue='grupo',
#             height=7, data=defunciones)
#c.fig.suptitle("Defunciones a causa de cáncer de mama por etapa de vida")
#plt.show()

#sns.set_theme( style='whitegrid')
#d=sns.relplot(x="edad", y="c530-c55Xm", size="c530-c55Xm",
#             sizes=(40,200), alpha=.5, palette="muted",hue='grupo',
#             height=7, data=defunciones)
#d.fig.suptitle("Defunciones a causa de cáncer cervicouterino por etapa de
vida")
#plt.show()

```

```

#sns.set_theme( style='whitegrid')
#d=sns.relplot(x="edad", y="c901-c959m", size="c901-c959m",
#             sizes=(40,200), alpha=.5, palette="muted",hue='grupo',
#             height=7, data=defunciones)
#d.fig.suptitle("Defunciones a causa de leucemia por etapa de vida")
#plt.show()

#
#c220_c959m13_18ConAcumularSinEdad = pd.read_csv('c220-c959m13-
18ConAcumularSinEdad.csv')
#sns.pairplot(c220_c959m13_18ConAcumularSinEdad, hue="grupo")
#plt.show()

#Mapa de calor
#defun = c220_c959m13_18ConAcumular.pivot("grupo", "edad", "c220-c249m")
# Draw a heatmap with the numeric values in each cell
#f, ax = plt.subplots(figsize=(150,150))
#sns.heatmap(defun, annot=True, linewidths=5, ax=ax)
#plt.show()

```