Universidad Nacional Autónoma de México

Facultad de Ciencias

Estimación del costo directo y costo indirecto asociado al número esperado de casos de cáncer en México al año 2020.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Daniel Antonio Armas Texta

TUTOR

M. en C. Federico Lasa Gonsebatt

Ciudad universitaria, CDMX **2017**

1. Datos del Alumno
   Armas
   Texta
   Daniel Antonio
   55 82 33 27
   Universidad Nacional Autónoma de
   México
   Facultad de Ciencias
   Actuaría
   410043916

2. Datos del tutor
   M en C
   Federico
   Lasa
   Gonsebatt

3. Datos del sinodal 1
   Dra
   María Cristina
   Gutiérrez
   Delgado

4. Datos del sinodal 2
   Dr
   Alejandro
   Mohar
   Betancourt

5. Datos del sinodal 3
   Act
   Ángel Manuel
   Godoy
   Aguilar

6. Datos del sinodal 4
   Dra
   Nancy
   Reynoso
   Noverón

7. Datos del trabajo escrito
   Estimación del costo directo y costo indirecto asociado al número esperado de casos de cáncer en México al año 2020
   100 p
   2017

Dedicado a mis padres

Índice general

Introducción.

El cáncer es una enfermedad de trascendencia a nivel internacional conocida principalmente por su agresividad hacia el cuerpo del paciente. Es una de las principales causas de muerte en el mundo y sin duda tema central de estudio para un gran número de investigadores.

En México durante el 2011, los principales tumores malignos por los cuales fallece la población con 20 años y más son[1]:

- En mujeres:
    1. Cáncer de mama (13.8 %).
    2. Cáncer cervicouterino (10.4 %).
    3. Cáncer de estómago (7.0 %).
- En hombres:
    1. Cáncer de próstata (16.9 %).
    2. Cáncer de bronquios y pulmón (12.8 %).
    3. Cáncer estómago (8.6 %).

La magnitud de las defunciones por tumores malignos aumenta conforme avanza la edad. Se observa que la tasa de mortalidad se incrementa de forma importante a partir de los 50 años y que a su vez, son menores en comparación con las tasas de mortalidad para la población de 80 años y más; por ejemplo en México durante el año 2011, 9 de cada 100 mil adultos entre 20 a 29 años de edad fallecian a consecuencia de algún tumor maligno mientras que 878 de cada 100 mil adultos de 80 años y más fallecian a consecuencia de la misma causa.

En México se hacen grandes esfuerzos por combatir dicha enfermedad, como la construcción de un nuevo edificio para el Instituto Nacional de Cancerología (INCan) así como la implementación del Seguro Popular que contempla el tratamiento algunas de las neoplasias importantes. Sin embargo, parece que nos enfretamos con

[1]Calculos propios a partir de datos de mortalidad del Intituto Nacional de Estadística y Geografía (INEGI).

un enemigo invisible ya que a pesar de que existe una gran variedad de información en todo el mundo, México no cuenta con un *Registro Nacional de Cáncer* o estadísticas propias para todas las variantes de dicha enfermedad.

Sin embargo, para algunos tipos de cáncer muy trascendentes (cáncer de mama, por ejemplo) se lleva un registro de casos nuevos, con lo cual se pueden estimar tasas de incidencia y así saber cuantos casos más se esperan en un periodo determinado de tiempo. Desafortunadamente, son muy pocas las neoplasias de las cuales se tiene tipo de seguimientos. En consecuencia surge la primer pregunta *hoy en día ¿cuántos enfermos de cáncer hay en el país?.*

Como se mencionó anteriormente, el cáncer es una enfermedad reconocida por la agresividad que tiene hacia el cuerpo del paciente, lo cual conlleva una demanda especial de atención y en consecuencia, un costo elevado para su tratamiento. Por ser una enfermedad a la cual, toda la población está en riesgo de desarrollar, el cáncer podría convertirse en un problema social. Cada año se destinan recursos al Sector Salud (350 mil millones de pesos durante el año 2014[2]) dirigidos principalmente al tratamiento de enfermedades. Por tanto, teniendo estos datos en mente y una vez contestada la pregunta anterior, otra pregunta que surge de manera natural en un contexto económico, social y epidemiológico es: *¿El país cuenta con suficientes recursos para atender a todos los enfermos de cáncer?* y por último *¿qué podemos hacer para minimizar la probabilidad de desarrollar cáncer?.*

El objetivo principal de este trabajo es responder las tres preguntas anteriores basados en modelos estadísticos y matemáticos, así como también con la ayuda de datos y estudios nacionales, información de organismos internacionales y otros estudios importantes sobre el tema. Para su mejor comprensión, el presente está organizado en 5 capítulos:

En el capítulo 1 se profundizará sobre el cáncer como enfermedad y la trascendencia que tiene a nivel internacional.

En el capítulo 2 se explorarán los modelos estadísticos más importantes para hacer predicciones y pronósticos, así como su construcción, implementación y verificación.

En el capítulo 3 se aplicarán las herramientas mencionadas en el capítulo 2 a datos relacionados con el cáncer complementadas con nociones y herramientas económicas, demográficas y actuariales para poder responder las preguntas planteadas.

En los capítulos 4 y 5 se presentarán los resultados del trabajo así como las conlcusiones del mismo.

[2]Según el Presupuesto de Egresos de la Federación para el Ejercicio Fiscal 2014 *Diario Oficial de la Federación 03-12-2013.*

Capítulo 1

El Cáncer, un enfoque epidemiológico.

1.1. El cáncer como enfermedad.

El cáncer es el nombre común que recibe un conjunto de enfermedades relacionadas en las que se observa un proceso descontrolado en la división de las células del cuerpo. Puede comenzar de manera localizada y diseminarse a otros tejidos circundantes. En general conduce a la muerte del paciente si este no recibe tratamiento adecuado. Los más comunes son: de piel, pulmón, mama y colonrectal ([**16**]).

El cáncer puede empezar casi en cualquier lugar del cuerpo humano, el cual está formado de trillones de células. Normalmente, las células humanas crecen y se dividen para formar nuevas células a medida que el cuerpo las necesita. Cuando las células normales envejecen o se dañan, mueren, y células nuevas las remplazan.

Sin embargo, en el cáncer, este proceso ordenado se descontrola. A medida que las células se hacen más y más anormales, las células viejas o dañadas sobreviven cuando deberían morir, y células nuevas se forman cuando no son necesarias. Estas células adicionales pueden dividirse sin interrupción y pueden formar masas que se llaman tumores.

Muchos cánceres forman tumores sólidos, los cuales son masas de tejido. Los cánceres de la sangre, como las leucemias, en general no forman tumores sólidos.

Los tumores cancerosos son malignos, lo que significa que se pueden extender a los tejidos cercanos o los pueden invadir. Además, al crecer estos tumores, algunas células cancerosas pueden desprenderse y moverse a lugares distantes del cuerpo por medio del sistema circulatorio o del sistema linfático y formar nuevos tumores lejos del tumor original.

Al contrario de los tumores malignos, los tumores benignos no se extienden a los tejidos cercanos y no los invaden. Sin embargo, a veces los tumores benignos pueden ser bastante grandes. Al extirparse, generalmente no vuelven a crecer,

mientras que los tumores malignos sí vuelven a crecer algunas veces. Al contrario de la mayoría de los tumores benignos en otras partes del cuerpo, los tumores benignos de cerebro pueden poner la vida en peligro.

El cáncer puede afectar a personas de todas las edades, incluso a fetos, pero el riesgo de sufrir los más comunes se incrementa con la edad. El cáncer causa cerca del 13 % de todas las muertes. De acuerdo con la Sociedad Americana del Cáncer, 7.6 millones de personas murieron por esta enfermedad en el mundo durante el año 2010 ([**4**]).

1.2. ¿Qué es la epidemiología?

La epidemiología es el estudio de la distribución y los determinantes de estados o eventos (en particular de enfermedades) relacionados con la salud y la aplicación de esos estudios al control de enfermedades y otros problemas de salud. Hay diversos métodos para llevar a cabo investigaciones epidemiológicas: la vigilancia y los estudios descriptivos se pueden utilizar para analizar la distribución, y los estudios analíticos permiten analizar los factores determinantes.

En epidemiología se estudian y describen la salud y las enfermedades que se presentan en una determinada población, para lo cual se tienen en cuenta una serie de patrones de enfermedad, que se reducen a tres aspectos: tiempo, lugar y persona: el tiempo que tarda en surgir, la temporada del año en la que surge y los tiempos en los que es más frecuente; el lugar (la ciudad, la población, el país, el tipo de zona) en donde se han presentado los casos, y las personas más propensas a padecerla (niños, ancianos, etc., según el caso).

La epidemiología surgió del estudio de las epidemias de enfermedades infecciosas; de ahí su nombre. Ya en el siglo XX los estudios epidemiológicos se extendieron a las enfermedades y problemas de salud en general, analizados mediante diversos métodos, entre los cuales los de la demografía y la estadística son especialmente importantes.

Existen 3 variables importantes que estudia la epidemiología, las cuales son:

- Mortalidad. Número de muertes que se producen en un período determinado en una población específica.
- Incidencia. Número de casos nuevos que se producen en un período determinado de tiempo en una población específica.
- Prevalencia. Número de personas vivas que han sido diagnosticadas con alguna enfermedad en un periódo del tiempo.

1.3. La epidemiología del cáncer a nivel internacional.

El cáncer es la principal causa de muerte a escala mundial. Se le atribuyen 8.2 millones de defunciones ocurridas en todo el mundo en 2012. Los principales tipos de cáncer son los siguientes [**8**]:

1. Pulmonar (1,59 millones de defunciones);
2. Hepático (745 000 defunciones);
3. Gástrico (723 000 defunciones);
4. Colorrectal (694 000) defunciones;
5. Mamario (521 000 defunciones);
6. Cáncer de esófago (400 000 defunciones).

Más del 30 % de las defunciones por cáncer podrían evitarse modificando o evitando los principales factores de riesgo [**14**], tales como:

- El consumo de tabaco.
- El exceso de peso o la obesidad.
- Las dietas malsanas con un consumo insuficiente de frutas y hortalizas.
- La inactividad física.
- El consumo de bebidas alcohólicas.
- Las infecciones ocasionadas por el Virus del Papiloma Humano y Virus de la Hepatítis B (PHV y VHB por sus siglas en inglés).
- Radiaciones ionizantes y no ionizantes.
- La contaminación del aire de las ciudades.
- El humo generado en la vivienda por la quema de combustibles sólidos.

El consumo de tabaco es el factor de riesgo más importante, y es la causa de aproximadamente un 22 % de las muertes mundiales por cáncer en general, y de acerca el 70 % de las muertes mundiales por cáncer de pulmón. En muchos países de ingresos bajos, hasta un 20 % de las muertes por cáncer son debidas a infecciones por el Virus de Papiloma Humano o Virus de Hepatitis B.

Además, la Organización Mundial de la Salud reportó las siguientes datos [**14**]:

- Los tipos más frecuentes de cáncer son diferentes en el hombre y en la mujer.

    - En 2012, los cánceres diagnosticados con más frecuencia en el hombre fueron los de pulmón, próstata, colon y recto, estómago e hígado.
    - En la mujer fueron los de mama, colon y recto, pulmón, cuello uterino y estómago.

- Aproximadamente un 30 % de las muertes por cáncer se deben a cinco factores de riesgo asociados al estilo de vida y alimentación (índice de masa corporal elevado, consumo insuficiente de frutas y verduras, falta de actividad física y consumo de tabaco y alcohol) y, por lo tanto, pueden prevenirse.
- El 70 % de todas las muertes por cáncer registradas en 2012 se produjeron en en África, Asia, América Central y Sudamérica.

En 2013, la OMS puso en marcha el Plan de Acción Global para la Prevención y el Control de las Enfermedades No Transmisibles 2013-2020 que tiene como objetivo reducir la mortalidad prematura por el 25 % de cáncer, enfermedades cardiovasculares, diabetes y enfermedades respiratorias crónicas. Algunas de las metas de aplicación voluntaria son especialmente importantes para la prevención del cáncer, como la que propone reducir el consumo de tabaco en un 30 % entre 2014 y 2025.

La OMS y el Centro Internacional de Investigaciones sobre el Cáncer colaboran con otras organizaciones que forman parte del Equipo de Tareas Interinstitucional de las Naciones Unidas sobre la Prevención y el Control de las Enfermedades No Transmisibles y con otros asociados a el fin de:

- Aumentar el compromiso político con la prevención y el control del cáncer;
- Coordinar y llevar a cabo investigaciones sobre las causas del cáncer y los mecanismos de la carcinogénesis en el ser humano;
- Efectuar un seguimiento de la carga de cáncer (como parte de la labor de la Iniciativa Mundial sobre Registros Oncológicos);
- Elaborar estrategias científicas de prevención y control del cáncer;
- Generar y divulgar conocimientos para facilitar la aplicación de métodos de control del cáncer basados en datos científicos;
- Elaborar normas e instrumentos para orientar la planificación y la ejecución de las intervenciones de prevención, detección temprana, tratamiento y atención;
- Facilitar la formación de amplias redes mundiales, regionales y nacionales de asociados y expertos en el control del cáncer;
- Fortalecer los sistemas de salud locales y nacionales para que presten servicios asistenciales y curativos a los pacientes con cáncer;

- Prestar asistencia técnica para la transferencia rápida y eficaz de las prácticas óptimas a los países en desarrollo.

Los datos y acuerdos anteriormente citados pueden ser consultados a profundidad en la bibliografía *referencia:* **[8]** y **[14]**.

1.4. International Agency For Research On Cancer *(IARC)* y *The GLOBOCAN project.*

La Agencia Internacional para la Investigación sobre el Cáncer (IARC por sus siglas en inglés) es la agencia especializada en el cáncer de la Organización Mundial de la Salud (OMS) creada el 20 de Mayo de 1965. El objetivo de la IARC es promover la colaboración internacional de la investigación en cáncer.

Es una agencia interdisciplinaria, reuniendo habilidades en epidemiología, ciencias de laboratorio, bioestadística entre otras.

Una característica importante de la IARC es su experiencia en la coordinación de la investigación entre países y sus organizaciones ya que su papel como organización internacional independiente facilita esta actividad.

La IARC tiene un papel muy importante en la descripción de la carga de cáncer en todo el mundo, a través de la cooperación y la ayuda para obtener acceso a los registros de cáncer así como el seguimiento de las variaciones geográficas y tendencias en el tiempo. El proyecto más representativo de la IARC es conocido como *The GLOBOCAN project*

El objetivo principal de *The GLOBOCAN project* es proporcionar estimaciones actuales de la incidencia, mortalidad y prevalencia de los principales tipos de cáncer, a nivel nacional, para 184 países en el mundo.

Dicho proyecto divide al planeta en 23 regiones. México se encuentra en la región de "Centro América" junto con Guatemala, Belice, El salvador, Honduras, Nicaragua, Costa Rica y Panamá.

Todos los datos los obtiene de fuentes oficiales internas o externas. Como fuentes internas encontramos principalmente censos, estadísticas oficiales y encuestas. Así como algunos trabajos y publicaciones relacionadas. Si algún país no cuenta con los datos necesarios para realizar el estudio, se utiliza los datos de los países vecinos de la misma región donde se encuentra el primero (Centro América para el caso de México).

Como fuentes externas se encuentran las bases de datos y registros de organizaciones internacionales, como el Banco Mundial y la Organización Mundial de la Salud, así como algunos estudios específicos de otras regiones, como el CI5 (Cancer Incidence in Five Continents) o el proyecto EUROCARE II.

Las estimaciones que realiza *The GLOBOCAN project* son de la siguiente forma.

- Incidencia. Si el país cuenta con datos sobre la incidencia por algún tipo específico de cáncer (o todos) la estimación se hace mediante un modelo de regresión lineal, en caso contrario se usan las estimaciones realizadas para la región a la cual pertenezca el país.

- Mortalidad. Para realizar las estimaciones del índice de mortalidad se usa el mismo procedimiento que para los casos de incidencia, con la salvedad que a este se le multiplica un factor de corrección para cuantificar un posible subregistro. En caso de que el país no cuente con los datos, la tasa será el promedio de las tasas de los países vecinos de la misma región.

- Prevalencia. Las estimaciones para la prevalencia se obtienen combinando el número anual de nuevos casos y la correspondiente tasa de mortalidad para dichos cánceres, ambos obtenidos de los puntos anteiores.

Este tipo de proyectos son de gran importancia para países en vías de desarrollo como México, ya que desafortunadamente no se cuenta con un registro nacional de cáncer o estudios similares. En consecuencia, todos los estudios epidemiológicos que se han hecho en materia de cáncer son con base a la experiencia propia de alguna dependencia de gobierno, o bien utilizando datos de estudios internacionales como *The GLOBOCAN project*.

Por otro lado, las estadísticas vitales de México (captadas y publicadas por el Instituto Nacional de Estadística y Geografía) son reconocidas a nivel internacinoal por organismos como la ONU y la OMS, lo cual es de gran ayuda en estudios epidemiológicos, ya que de dichos datos podemos obtener información sobre nacimientos, defunciones, migración, emigración, matrimonios, divorcios, entre muchas otras variables.

Capítulo 2

Modelos estadísticos para predicciones y pronósticos.

Modelar y pronósticar son las dos prioridades de la mayoría de los investigadores y temas centrales de estudio para la estadística. En este capítulo se abordaran las nociones principales y básicas para las dos herramientas más frecuentemente usadas que son: El análisis de regresión y Modelos de series de tiempo.

2.1. Análisis de Regresión.

El análisis de regresión es una de las herramientas estadísticas más frecuentemente usadas para analizar datos multivariados ya que se usa para investigar y modelar relaciones entre variables. Su gran atractivo conceptual reside en utilizar una ecuación simple para expresar la relacion entre un conjunto de variables. Así mismo, el analisis de regresión es teóricamente interesante por el nivel matemático requerido para su comprensión. Para ocupar de manera adecuada dicha herramienta se necesita una amplia apreciación tanto de la teoría matemática como del contexto de aplicación de la misma ya que siempre se debe de dar la interpretación del resultado, y claramente, este último debe de tener sentido dentro del contexto del problema.

Este capítulo está basado en el libro *Introduction to Linear Regression Analysis* ([**2**]).

2.1.1. Modelo de regresión lineal simple y estimación de parámetros.

***Definición* 2.1** *El modelo de regresión lineal simple está dado por:*

$$y = \beta_0 + \beta_1 x + \varepsilon$$

donde:

y es la variable de respuesta
x es la varibale regresora
β_0 es la ordenada al origen
β_1 es la pendiente de la recta y
ε es el componente de error

Se asume que ε es una variable aleatoria de media cero y varianza σ^2. Así mismo, si se tiene más de una observación, se supone que un error no depende del valor de algún otro error, i.e., los errores son *no correlacionados*.

Cabe resaltar que x son observaciones y y es la respuesta correspondiente a cada valor de x.

En consecuencia, del modelo de regresión lineal simple se busca obtener los estimadores de β_0 y β_1 ($\hat{\beta}_0$ y $\hat{\beta}_1$, respectivamente) para así obtener la *recta de regresión ajustada* dada por:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

de tal forma que las observaciones estimadas ($\hat{y}$) estén lo más cerca ('se parezcan') a las observaciones reales (y).

Para poder obtener los estimadores de β_0 y β_1 se ocupará el método de **mínimos cuadrados**, es decir, se busca a $\hat{\beta}_0$ y $\hat{\beta}_1$ tales que la suma de los cuadrados de la diferencia entre y y $\hat{y}$ sea mínima.

Sean (x_i, y_i) con $i = 1, 2, ..., n$ parejas de datos, entonces el modelo de regresión se puede escribir como:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, ..., n$$

Definamos a $S(\beta_0, \beta_1)$ como:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x_i)^2$$

Entonces, los estimadores por mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ deben satisfacer que:

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i) = 0$$

y

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i) x_i = 0$$

Si despejamos a $\hat{\beta}_0$ y $\hat{\beta}_1$ de las ecuaciones anteriores obtenemos el siguiente sistema:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=i}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$

a este último sistema de ecuaciones se le conoce como *ecuaciones normales*, cuya solución esta dada por:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

y

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=i}^{n} x_i^2 - \frac{\left(\sum_{i=i}^{n} x_i\right)^2}{n}}$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

y

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Además de estimar a β_0 y β_1 necesitamos estimar a σ^2 para poder calcular intervalos de confianza y hacer pruebas de hipótesis.

***Definición* 2.2** . *El i-ésimo residual (e_i) se define como*

$$e_i = y_i - \hat{y}_i$$

Cuando no se tiene información alguna de σ^2 se estima a partir de las sumas de cuadrados de los errores, o bien, la suma de cuadrados de los *residuales*.

Entonces buscamos estimar a σ^2 a partir de

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Notemos que la suma de cuadrados de los residuales tiene $(n-2)$ grados de libertad, ya que los otros dos grados de libertad estan asociados a la estimación de $\hat{\beta}_0$ y $\hat{\beta}_1$ que son necesarios para calcular a $\hat{y}_i$

Por tanto, el estimador insesgado para σ^2 está dado por:

$$MS_E = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}$$

A $\hat{\sigma}^2$ se le conoce como media cuadrada del error o *media cuadrada residual*. A la raíz cuadrada de $\hat{\sigma}^2$ se le conoce como *error estándar de la regresión* y está en las mismas unidades que la variable de respuesta y

2.1.2. Pruebas de hipótesis.

Uno de los temas más relevantes en la estadística son las pruebas de hipótesis ya que es de importancia saber si existe evidencia estadística para pensar que las hipótesis del modelo estan siendo violadas.

De ahora en adelante supondremos que $\{\varepsilon_i\}_{i=1}^n$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas $N(0,\sigma^2)$ y la notación será:

$$\{\varepsilon_i\}_{i=1}^n \sim iidN(0,\sigma^2)$$

Ahora supóngase que se desea probar la hipótesis de que la ordenada al origen (intercepto) es igual a una constante, digamos β_{00}. Las hipótesis están dadas por:

$$H_0 : \beta_0 = \beta_{00}$$

$$vs$$

$$H_1 : \beta_0 \neq \beta_{00}$$

La cual es una prueba de dos colas.

Dado que $\varepsilon_i \sim N(0,\sigma^2)$ entonces $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.
Por otro lado, $\hat{\beta_0}$ es combinación lineal de las observaciones, asi que
$\hat{\beta_0} \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{\sum_{i=i}^n x_i^2 - \frac{\left(\sum_{i=i}^n x_i\right)^2}{n}}\right)\right)$. Por tanto, la estadística de

$$Z_0 = \frac{\hat{\beta_0} - \beta_{00}}{\sqrt{\sigma^2(\frac{1}{n} + \frac{\bar{x}}{\sum_{i=i}^n x_i^2 - \frac{\left(\sum_{i=i}^n x_i\right)^2}{n}})}}$$

Se distribuye $N(0,1)$, si suponemos H_0 cierta. Ahora el problema es conocer a σ^2, pero recordemos que MS_E es un estimador insesgado para σ^2 y además $\frac{(n-2)MS_E}{\sigma^2}$ es una variable aleatoria χ^2_{n-2}. Por último, sabemos que MS_E y β_0 variables aleatorias independientes.

Todo lo anterior implica que si sustituimos σ^2 por MS_E en Z_0 entonces la estadística

$$t_0 = \frac{\hat{\beta_0} - \beta_{00}}{\sqrt{MS_E(\frac{1}{n} + \frac{\bar{x}}{\sum_{i=i}^n x_i^2 - \frac{\left(\sum_{i=i}^n x_i\right)^2}{n}})}}$$

se distribuye como una $t - student$ con $n-2$ grados de libertad bajo H_0. Por lo tanto, la regla de decision será: *rechazar H_0 a un nivel de significancia α si:*

$$|t_0| > t_{\frac{\alpha}{2}}^{n-2}$$

Se puede hacer de manera análoga para las hipótesis

$$H_0 : \beta_1 = \beta_{10} \qquad vs \qquad H_1 : \beta_1 \neq \beta_{10}$$

Pero hay un caso que es de particular interés, y es cuando $\beta_{10} = 0$, es decir:

$$H_0 : \beta_1 = 0$$

$$vs$$

$$H_1 : \beta_1 \neq 0$$

Esta prueba esta relacionada con la *significancia de la regresión* ya que si $\beta_1 = 0$ implicaría cualquiera de las siguientes dos casos:
1. La variación de los valores de x perturban de manera poco significatica a y, entonces el modelo lineal sería una linea horizontal donde el mejor estimador para cualquier valor de x seria $\bar{y}$.
2. La relación entre x y y no es lineal.

En cualquiera de los dos casos diremos que el modelo no es significativo.

Para abordar este problema utilizaremos la tabla de análisis de varianza (conocida como *tabla ANOVA*) la cual tiene la siguiente estructura

Fuente de variación	*Suma de cuadrados*	*Grados de libertad*	*Cuadrados medios*	F_0
Regresión	$SS_R = \hat{\beta}_1 S_{xy}$	1	$MS_R = \frac{SS_R}{1}$	$\frac{MS_R}{MS_E}$
Residuales	$SS_E = S_{yy} - \hat{\beta}_1 S_{xy}$	$n-2$	$MS_E = \frac{SS_E}{n-2}$	
Total	S_{yy}	$n-1$		

donde:
$SS_{xy} = \sum_{i=1}^{n} x_i y_i$
$SS_{yy} = \sum_{i=i}^{n} y_i^2 - \frac{\left(\sum_{i=i}^{n} y_i\right)^2}{n}$

Si suponemos H_0 cierta, entonces F_0 se distribuye $F_{(1,n-2)}$.

Por lo tanto, la regla de decisión está dada por: *rechazar H_0 a un nivel de significancia α si*:

$$F_0 > F^{\alpha}_{(1,n-2)}$$

Si no se rechaza H_0, diremos que *la regresión no es significativa* o bien que *el modelo lineal es no significativo*

Además, de la tabla de análisis de varianza podemos obtener el *coeficiente de determinación* (R^2), el cual se cualcula de la sigueinte forma:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}}$$

Notemos que S_{yy} es la medida de variabilidad de y sin considerar el efecto de la variable regresora x y SS_E es la medida de variabilidad restante de y después de que x fue considerada. Por lo anterior, a R^2 se le interpreta como la proporcion de la variabilidad explicada por la variable regresora x.

Como $0 \leq SS_E \leq Syy \Rightarrow 0 \leq R^2 \leq 1$. Entonces, valores cercanos a 1 implican que la mayor parte de la variabilidad de y esta siendo explicada por el modelo de regresión.

Sin embargo, la etadística R^2 debe de usarse con cuidado, ya que siempre es posible hacerla más grande, agregando un número suficiente de términos al modelo. Por ejemplo, si no hay observaciones repetidas, (a un mismo nivel de x correspondan dos valores diferentes de y) entonces un polinomio de grado $n-1$ se ajustará perfectamente ($R^2 = 1$) a un conjunto de n observaciones.

Por otro lado, R^2 también puede incrementar agregando variables regresoras al modelo, lo cual no implica que el nuevo modelo sea mejor que el anterior. Salvo que la suma de los cuadrados del error sea menor respecto al primero modelo, el nuevo modelo tendrá una media de cuadrados del error (MS_E) mayor que el original, esto se debe a que se pierde un grado de libertad al estimar una segunda variable regresora. En consecuancia, el segundo modelo puede ser peor que el primero.

En 1973, Hann ([**2**]) observó que el valor esperado de R^2, proveniente de una recta de regresión ajustada, se aproxima:

$$\mathbb{E}(R^2) \simeq \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\beta}_1^2 S_{xx} + \sigma^2}$$

Claramente el valor esperado de R^2 aumenta (o disminuye) conforme S_{xx}, (la cual es una medida de variación de la variable regresora x) aumenta (o disminuye). Por lo tanto, un valor grande de R^2 puede ser el resultado de que x haya variado un rango demasiado amplio. Por otro lado, un valor pequeño de R^2 puede ser el resultado de que el rango de variación de x fue demasiado pequeño para permitir una relación con y.

2.1.3. Predicción de nuevas observaciones.

Una aplicación importante para el modelo de regresión es la predicción de nuevas observaciones de y correspondientes a un nivel específico de x. Si x_0 es el valor de interés de la variable regresora, entonces:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

es el valor estimado del nuevo valor de respuesta y_0

Construyamos el *intervalo de predicción.*

Notemos que:

$$\psi = y_0 - \hat{y_0}$$

es una variable aleatoria normal con media cero y varianza

$$V(\psi) = V(y_0 - \hat{y_0}) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=i}^{n} x_i^2 - \frac{\left(\sum_{i=i}^{n} x_i\right)^2}{n}} \right]$$

ya que la observación futura y_0 es independiente de $\hat{y}_0$. Si usamos $\hat{y}_0$ para predecir a y_0, entonces el error estándar de $\psi = y_0 - \hat{y}_0$ es una estadística adecuada para construir el intervalo de predicción.

Por lo tanto, el intervalo de $100(1-\alpha)\,\%$ para la predicción está dado por:

$$y_0 \in \left[\hat{y}_0 \ \begin{matrix} + \\ - \end{matrix} \ t_{\frac{\alpha}{2}}^{n-2} \sqrt{MS_E \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=i}^{n} x_i^2 - \frac{\left(\sum_{i=i}^{n} x_i\right)^2}{n}} \right)} \right]$$

Notemos que el intervalo para la predicción es mínimo cuando $x_0 = \bar{x}$ y aumenta conforme $|x_0 - \bar{x}|$ crece.

2.2. Modelos de series de tiempo.

Una serie de tiempo es una secuencia de datos estadísticos ordenadas cronológicamente y espaciados de manera equidistante, así los datos usualmente son dependientes entre sí. Los modelos para su estudio son complejos pero su atractivo principal radica en ser mejores modelos para datos que presentan alguna tendencia, no necesariamente lineal, a lo largo del tiempo.

Este capitulo está basado en el libro de *Introduction to Time Series and Forecasting* y *Análisis Estadístico de Series de Tiempo Económicas.* (*Bibliografía:* **[1]** y **[3]**).

2.2.1. Modelos ARIMA para series de tiempo univariadas.

***Definición* 2.3** *Una serie de tiempo es un conjunto de observaciones X_t cada una de las cuales se registra a un tiempo específico.*

***Definición* 2.4** *Un modelo de series de tiempo para los datos observados es una especificación de las distribuciones (o posiblemente sólo en terminos de sus medias y covarianzas) de una sucesión de variables aleatorias $\{X_t\}$ de las cuales supondremos que las observaciones previas son una realización.*

***Definición* 2.5** *El ruido blanco es una colección de variables aleatorias no correlacionadas con media cero y varianza finita. Se denotará de la siguiente forma*

$$Z_t \sim WN(0, \sigma_Z^2)$$

Sea $\{X_t\}$ una serie de tiempo con segundo momento finito, entonces la función de medias se define como:

$$\mu_X(t) := \mathbb{E}(X_t)$$

y la función de autocovarianza (ACVF) se define cómo

$$\gamma_X(r, s) := cov(X_r, X_s) \qquad \forall\ s, r \in \mathbb{Z}$$

***Definición* 2.6** *Sea $\{X_t\}$ serie de tiempo y sea $F_X\left(x_{t_1+n}, ..., x_{t_k+n}\right)$ su función de distribución. Se dice que $\{X_t\}$ es estacionaria fuertemente si $\forall\ k, n \in \mathbb{N}$ y $\forall\ t_1, ..., t_k$ se tiene:*

$$F_X\left(x_{t_1+n}, ..., x_{t_k+n}\right) = F_X\left(x_{t_1}, ..., x_{t_k}\right)$$

Es decir, F_X no es función del tiempo.

***Definición* 2.7** *La serie de tiempo $\{X_t\}$ se dice que es estacionaria debilmente si:*

1. $\mathbb{E}[X_t^2] < \infty \quad \forall t \in T$

2. $\mathbb{E}[X_t] = \mu_X(t) = m \quad con \quad m = cte \quad \forall\ t \in T$

3. $\gamma_X(r,s) = \gamma_X(r+t, s+t) \qquad \forall\ r,s,t \in T$

A partir de aquí, cada vez que se hable de *estacionariedad*, se hará referencia a *estacionariedad débil*

***Definición* 2.8** *Si $\{X_t\}$ es una serie de tiempo estacionaria, entonces la ACVF se define como:*

$$\gamma_X(h) := cov(X_t, X_{t+h})$$

Además, se define a la función de autocorrelación (ACF) en el rezago h como:

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}$$

***Definición* 2.9** *La función de autocorrelacion parcial (PACF) de un proceso estacionario $\{X_t\}$, es la función $\alpha(h)$ dada por:*

$$\alpha(h) = \begin{cases} 1 & si \quad h = 0 \\ \phi_{hh} & si \quad h \geq 1 \end{cases}$$

donde ϕ_{hh} es la última componente de la matriz

$$\phi_h = \Gamma_h^{-1}\gamma_h$$

donde:
$\Gamma_h^{-1} := [\gamma(i-j)]_{i,j=1}^h$ *es la matriz de varianzas y covarianzas de* $\{X_t\}$
$\gamma_h = [\gamma(1), \gamma(2), ..., \gamma(h)]'$

Definición* 2.10** *El operador B, conocido como* ***operador de retraso*, se define como:*

$$BX_t = X_{t-1}$$

Definición* 2.11** *El operador ∇, conocido como* ***operador de primera diferencia*, se define como:*

$$\begin{aligned} \nabla X_t &= (1-B)X_t \\ &= X_t - X_{t-1} \end{aligned}$$

Observación: Los polinomios en B y ∇ tienen la cualidad de poderse operar como polinomios de variable real.

De ahora en adelante, cuando se haga referencia a una serie de tiempo estacionaria, supondremos que esta tiene de media cero. Este supuesno no afecta ya que dada una serie de tiempo estacionaria $\{X_t\}$, si $\mathbb{E}(X_t) = \mu$ entonces podemos definir una nueva serie de tiempo $\{\dot{X}_t\}$ de la siguiente manera:

$$\dot{X}_t := X_t - \mu$$

Por tanto, la nueva serie $\{\dot{X}_t\}$ seguirá siendo un proceso estacionario pero con media cero.

***Definición* 2.12** *Un modelo auntoregresivo de orden p (AR(p)) está definido por la siguiente estructura:*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + Z_t$$

donde:
$\{X_t\}$ *es un proceso estacionario*
$\phi_1, ..., \phi_p$ *son constantes tales que* $\phi_p \neq 0$
$\{Z_t\} \sim WN(0, \sigma_Z^2)$

***Definición* 2.13** *El polinomio auntorregresivo de orden p se define como:*

$$\phi(B) := 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$$

Dada la definición anterior podemos reescribir un modelo auntorregresivo de orden p como:

$$\phi(B) X_t = Z_t$$

***Definición* 2.14** *El modelo de promedios móviles de orden q (MA(q)) está definido por la siguiente estructura:*

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q}$$

donde:
$\theta_1, ..., \theta_q$ *son constantes tales que* $\theta_q \neq 0$
$\{Z_t\} \sim WN(0, \sigma_Z^2)$

***Definición* 2.15** *El operdador o poliniomio de promedios móviles de orden q se define como:*

$$\theta(B) := 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q$$

Dada la definición anterior, un modelo de promedios móviles se puede escribir cómo:

$$X_t = \theta(B) Z_t$$

***Definición* 2.16** *Un modelo autorregresivo de promedios móviles de orden p y q (ARMA(p, q)) se define cómo:*

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + ... + \theta_q Z_{t-q}$$

donde:
$\{X_t\}$ *es un proceso estacionario*
$\{Z_t\} \sim WN(0, \sigma_Z^2)$
además, el polinomio autorregresivo no tiene raíces en común con el polinomio de promedios móviles

En termino de los polinomios autorregresivos y de promedios móviles, un modelo $ARMA(p,q)$ se puede escribir como:

$$\phi(B)X_t = \theta(B)Z_t$$

***Definición* 2.17** *Decimos que un proceso $\{X_t\}$ es **causal (o función causal de $\{Z_t\}$)** si X_t puede ser expresado en términos del valor actual y pasados de Z_s con $s \leq t$*

***Definición* 2.18** *Decimos que un proceso $\{Z_t\}$ es **invertible (o función invertible de $\{X_t\}$)** si Z_t puede ser expresado en términos del valor actual y pasados de X_s con $s \leq t$*

***Resultado* 2.1** *Un proceso $ARMA(p,q)$ es causal si existen constantes $\{\psi_j\}$ tales que:*

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

Verificar la condicion de causalidad para un proceso $ARMA(p,q)$ es equivalente a verificar:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p \neq 0 \quad \forall\ |z| \leq 1, \quad z \in \mathbb{C}$$

***Resultado* 2.2** *Un proceso $ARMA(p,q)$ es invertible si existen constantes $\{\pi_j\}$ tales que:*

$$\sum_{j=0}^{\infty} |\pi_j| < \infty$$

y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

Verificar la condicion de causalidad para un proceso $ARMA(p,q)$ es equivalente a verificar:

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q \neq 0 \quad \forall\ |z| \leq 1, \quad z \in \mathbb{C}$$

Propiedades de los procesos de promedios móviles de orden q ($MA(q)$):

- Siempre son invertibles.
- Su ACF es cero después del **lag** (retraso) q.

- Su $PACF$ decrese suavemente.

Propiedades de los procesos de autorregresivos de orden p ($AR(p)$):

- Siempre son causales.
- Su ACF decrese suavemente.
- Su $PACF$ es cero después del **lag** (retraso) p.

A diferencia de de los $AR(p)$ y $MA(q)$, los $ARMA$ no siempre son causales o invertibles y tanto su ACF como su $PACF$ decresen suavemente.

Definición* 2.19** ***(Criterio del AICC) *Escoger $p, q, \hat{\bar{\phi}}_p$ y $\hat{\bar{\theta}}_q$ tales que se minimice el $AICC$ definido por:*

$$AICC = -2lnL\left(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q, \hat{\sigma}^2\right) + \frac{2(p+q+1)n}{n-p-q-2}$$

donde: $L\left(\bar{\phi}_p, \bar{\theta}_q, \sigma^2\right)$ es la función de Verosimilitud.

***Definición* 2.20** *Si d es un entero no negativo, entonces $\{X_t\}$ es un proceso $ARIMA(p,d,q)$ si:*

$$Y_t := (1-B)^d X_t = \nabla^d X_t$$

es un proceso $ARMA(p,q)$ causal.

Los procesos $ARIMA(p,d,q)$ son una generalización de los procesos $ARMA(p,q)$ que se usan cuando $\{X_t\}$ no es un proceso estacionario.

2.2.2. Estimación de parámetros.

La etapa de estimación presupone que se ha identificado el modelo y que, de ser éste adecuado, lo único que resta es encontrar los mejores valores de los parámetros para que dicho modelo represente apropiadamente a la serie considerada. La estimación podría haerse de forma arbitraria pero evidentemente es preferible utilizar un método objetivo y estadísticamente apropiado. El método mas usado para estimar los parámetros del modelo por máxima verosimilitud ([**3**]).

A partir de este punto supondremos que $\{Z_t\}$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas $N(0, \sigma_Z^2)$ y diremos que $\{X_t\}$ es una serie de tiempo con ruido *gaussiano*, o bien, una serie *gaussiana*.

Sea $\{X_t\}$ un proceso $ARIMA(p, d, q)$, causal e invertible, con ruido *gaussiano* con n observaciones, entonces la densidad conjunta de los errores aleatorios esta dada por:

$$f(Z_{d+p+1}, ..., Z_n) = (2\pi)^{-(\frac{n-d-p}{2})}\sigma_Z^{-n+d+p} \exp\{-\sum_{t=d+p+1}^{n} \frac{Z_t^2}{2\sigma_Z^2}\}$$

Cómo $\{X_t\}$ un proceso $ARIMA(p, d, q)$, entonces:

$$Z_t = X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}$$

lo cual nos permite obtener la función de densidad conjunta de $X_{d+p+1}, ..., X_n$ como:

$$f(X_{d+p+1}, ..., X_n) = f(Z_{d+p+1}, ..., Z_n) \prod_{t=d+p+1}^{n} \left|\frac{dX_t}{dZ_t}\right|$$

$$= (2\pi)^{-(\frac{n-d-p}{2})}\sigma_Z^{-n+d+p} \exp\left\{-S(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2)\right\}$$

donde:
$S(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2) = \sum_{t=d+p+1}^{n} \frac{(X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q})^2}{2\sigma_Z^2}$

La cual permite calcular probabilidad de la distribución normal multivariada una vez que se conocen a los parámetros $\bar{\phi_p} = (\phi_1, ..., \phi_p)'$, $\bar{\theta_q} = (\theta_1, ..., \theta_q)'$ y σ_Z^2; ahora bien, lo que se conoce en realidad es $(X_{d+p+1}, ..., X_n)'$ y lo que se desconoce es ϕ_p, θ_q y σ_Z^2, por lo cual, la función de verosimilitud con la que se trabajará será:

$$L(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2 | X_{d+p+1}, ..., X_n) = (2\pi)^{-(\frac{n-d-p}{2})}\sigma_Z^{-n+d+p} \exp\left\{-S(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2)\right\}$$

Para maximizar $L(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2 | X_{d+p+1}, ..., X_n)$, en primer lugar se eligen los valores de $\bar{\phi_p}$ y $\bar{\theta_q}$ de tal forma que minimicen a $S(\bar{\phi_p}, \bar{\theta_q}, \sigma_Z^2)$ y posteriormente se determina el estimador de σ_Z^2

Supóngase en principio, que ya se encontraron los valores $\hat{\bar{\phi}}_p$ y $\hat{\bar{\theta}}_q$ de tal forma que $S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)$ es mínimo, entonces se procede maximizar la función de verosimilitud con respecto a σ_Z^2. Para esto se utilizará la *log-verosimilitud*.

$$l(\sigma_Z^2|X_{d+p+1}, ..., X_n, \hat{\phi}_p, \hat{\theta}_q)) = -\frac{(n-d-p)}{2}[ln(2\pi) - ln(\sigma_Z^2)] - \frac{S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)}{2\sigma_Z^2}$$

Para lograr dicha maximización se debe encontrar $\hat{\sigma}_Z^2$ de tal forma que:

$$\left.\frac{\partial l}{\partial \sigma_Z^2}\right|_{\hat{\sigma}_Z^2} = \frac{-n+d+p}{2\hat{\sigma}_Z^2} + \frac{S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)}{2\hat{\sigma}_Z^4}$$

Por lo tanto, es estimador máximo verosimil de σ_Z^2 está dado por:

$$\hat{\sigma}_Z^2 = \frac{S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)}{n-d-p}$$

En la práctica prefiere usarse al estimador insesgado ya que considera la corrección por grados de libertad usados para estimar a todos los parámetros del modelo y se calcula cómo:

$$\hat{\sigma}_Z^2 = \frac{S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)}{n-d-p-q-1}$$

Entonces, el problema de maximizar a $L(\bar{\phi}_p, \bar{\theta}_q, \sigma_Z^2|X_{d+p+1}, ..., X_n)$ se reduce a minimizar $S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)$. Sin embargo, al buscar dichos valores, se llegan a ecuaciones no lineales sin solución analítica cerrada, por lo cual se deben utilizar métodos numéricos para encontrar el mínimo de $S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)$.

En su libro, Box y Jenkins (1970) sugieren un método de estimación no lineal para $\bar{\phi}_p$ y $\bar{\theta}_q$ basados en el algoritmo de Marquardt (1963), que es utilizado por varios softwares de cómputo estadístico y que permite obtener no solo las estimaciones puntuales de los parámetros, sino también intervalos de confianza. Dicho método tiene como fundamento un desarrollo en series de Taylor que linealiza a Z_t condicionada a que se conocen $X_{d+p+1}, ..., X_n$ y los valores iniciales de los parámetros $\bar{\phi}_p^*$ y $\bar{\theta}_q^*$. Los valores iniciales serán corregidos iterativamente con el objetivo final de minimizar $S(\hat{\bar{\phi}}_p, \hat{\bar{\theta}}_q)$. Aunque el proceso de estimación puede ser sensible a los valores iniciales, estos ya son dados autómaticamente por la mayoría de los softwares de cómputo estadístico.

2.2.3. Verificación de supuestos.

Una de las formas más claras y simples para detectar violaciones a los supuestos de los modelos es a través del análisis de residuales ([**3**]). Recordemos que los residuales se definen como la diferencia entre los valores observados y los valores estimados por el modelo:

$$\hat{Z}_t = X_t - \hat{X}_t$$

Los supuestos a verificar son los siguientes:

1. $\{Z_t\}$ tiene media cero. Para verificar este supuesto, se debe de calcular la media arítmetica ($\hat{\bar{Z}}$) y la desviación estándar de los residuales ($\hat{\sigma}^2_Z$). Si:

$$\left| \frac{\sqrt{n-d-p}\ \hat{\bar{Z}}}{\hat{\sigma}^2_Z} \right| < \Phi_{1-\frac{\alpha}{2}}$$

   donde $\Phi_{1-\frac{\alpha}{2}}$ es el cuantil $(1-\frac{\alpha}{2})$ de una distribución normal estándar.

   En este caso, se concluye que un nivel de significancia α no existe evidencia de que la media del proceso de ruido blanco sea distinta de cero, por tanto no se rechaza el supuesto.

2. $\{Z_t\}$ tiene varianza constante. La verificación de este supuesto se hace mediante la gráfica de los residuales contra el tiempo para observar visualmente si la varianza parece ser o no constante. O bien se puede utilizar la prueba de Bartlett para homogeneidad de varianzas.

3. Las variables aleatorias $\{Z_t\}$ son mutuamente no correlacionadas. Para la verificación se usa la prueba de Ljung-Box.

4. Z_t tiene distribución normal. Para verificar esta prueba se usa la prueba de Lilliefors para normalidad.

Además, un supuesto extra es el de pensar que el modelo es **parsimonioso**, es decir, no se puede reducir el número de parámetros en el modelo ya que todos son necesarios para explicar el comportamiento del modelo. Lo que se hace para verificar este supuesto es construir intervalos de confianza para los parámetros θ con la sigiuente estructura:

$$\left(\hat{\theta} - \Phi_{\frac{\alpha}{2}} \sqrt{\hat{Var}(\theta)}, \hat{\theta} + \Phi_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}(\theta)} \right)$$

En caso de que el intervalo contenga al 0, entonces ese parámetro será no significativo y se deberá de omitir del modelo. En consecuencia se deberan de reestimar los parámetros omitiendo los parámetros no significativos.

2.2.4. Pronóstico e intervalo de confianza.

Uno de los fines mas frecuentes al contruir un modelo para una serie de tiempo dada es el pronóstico([**1**] **y** [**3**]).

***Definición* 2.21** *El mejor estimador lineal (**BLP**) de X_{n+h} en términos de $(X_1, ..., X_n)$ se define como:*

$$P_n X_{n+h} = a_1 X_n + ... + a_n X_1$$

*Y es aquella combinación lineal que mejor aproxima a X_{n+h} minimizando el **error cuadrático medio**.*

***Resultado* 2.3** *Sea $\{X_t\}$ es un proceso $ARIMA(p, d, q)$ de tal forma que $Y_t = (1 - B)^d X_t$ es un proceso $ARMA(p, q)$ causal y el vector $(X_{1-d}, ..., X_0)$ es no correlacionado con Y_t $\forall t > 0$, entonces el mejor predictor lineal h-pasos hacia adelante para X_{n+h} está dado por:*

$$P_n X_{n+h} = \sum_{j=1}^{p+d} \phi_j^* P_n X_{n+h-j} + \sum_{j=h}^{q} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})$$

donde:
$\phi^*(z) = (1 - z)^d \phi(z) = 1 - \phi_1^* z - ... - \phi_{p+d}^* z^{p+d}$
$\hat{X}_{n+h-j} = P_n X_{n+h-j}$ *es el predictor a un paso.*
y los coeficientes $\theta_{n+h-1,h}, ..., \theta_{n+h-1,q}$ se calculan de las ecuaciones recursivas

$$\upsilon_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = \upsilon_k^{-1} \left(\kappa(n + 1, k + 1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta n, n - j \upsilon_j \right), \qquad 0 \leq k < n,$$

y

$$\upsilon_n = \kappa(n + 1, n + 1) - \textstyle\sum_{j=0}^{n-1} \theta_{n,n-j}^2 \upsilon_j.$$

con:

$$\kappa(i, j) := \mathbb{E}(X_i X_j)$$

Notemos que $P_n X_{n+1-j} = X_{n+1-j}$ para cada $j \geq 1$ ya que se trata de la información anteriormente observada.

Por último, si $\{Y_t\}$ es un proceso $ARMA(p, q)$ con ruido *gaussiano*, entonces para cada $h \geq 1$ se cumple que el error del pronóstico

$$(Y_{n+h} - P_n Y n + h) \sim N(0, \sigma_Z^2 \sum_{j=0}^{h-1} \psi_j)$$

por lo tanto, las bandas de $100(1-\alpha)\,\%$ de confianza para el prónostico estan dadas por:

$$Y_{n+h} \in \left[P_n Y_{n+h} \ \stackrel{+}{-} \ \Phi_{1-\frac{\alpha}{2}} \sqrt{\sigma_Z^2 \sum_{j=0}^{h-1} \psi_j} \right]$$

2.2.5. Metodología de Box-Jenkins.

Box y Jenkins en 1970 sugieren una metodología para trabajar con series de tiempo y que hoy en día es de los protocólos mas utilizados cuando de series de temporales se trata. A lo largo de este tema se ha seguido con los pasos de Box-Jenkins y acontinuación se resumirá.

1. **Identificación.** Representar gráficamente la serie, además de su función de autocorrelación simple (ACF) y función de autocorrelación parcial ($PACF$). La gráfica de la serie nos indica si la serie es estacionaria o no. Según los motivos por los que la serie no es estacionaria, tendremos que aplicar los siguientes procedimientos hasta hacerla estacionaria.

   - Si tiene tendencia: Tomaremos diferencias regulares hasta que desaparezca. Normalmente el orden de la diferencia es 1, y raramente será mayor a 3.
   - Si es heterocedástica, es decir, no tiene varianza constante, habrá que transformar la serie. Con tomar el logaritmo en muchos casos es suficiente, aunque existen algunas transformaciones más sofisticadas, como las de Box-Cox.

   Una vez que el gráfico de la nueva serie (transformación de la original) indica que es estacionaria, podemos intentar deducir la estructura de la serie (no la de la serie original) observando su ACF y $PACF$.

2. **Estimación.** Observando las dos gráficas del ACF y $PACF$ de la serie transformada podemos hacernos una idea del modelo que describe nuestra serie, o al menos de cuáles son los primeros candidatos que debemos probar. Para comprobar analíticamente (no visualmente) un modelo frecuentemente se ajusta varios modelos candidatos $ARIMA(p,d,q)$ y escogeremos como un buen modelo aquel que tenga los residuales semejantes al de un ruido blanco, además que tenga los valores del $AICC$ (Criterio de Información de Akaike corregido) menor con relación al resto de los modelos candidatos.

3. **Verificación de los supuestos.** En este punto es donde se verifican los supuestos explicados con anterioridad (**sección 2.4**).

4. **Pronóstico.** También conocido como uso del modelo. El modelo se ocupa para los fines que fue contruido. El éxito de los modelos $ARIMA$ radica en el pronóstico ya que son buenos para realizar predicciones a corto plazo.

Capítulo 3

Desarrollo del problema.

En este capítulo se presentan los modelos utilizados para estimar el costo directo e indirecto de los casos de cáncer. El desarrollo se dividirá en 2 temas importantes:

1. Estimación de casos prevalentes.

2. Estimación del costo total.

Nota: todas las anualidades fueron calculadas con base a la tabla de tasas de mortalidad de activos para la seguridad social 1997 (EMSSA-97) tanto para hombres como para mujeres.

3.1. Estimación de los casos prevalentes en cáncer en México al año 2020.

El conocimiento de la población de un país, desde el punto de vista de su composición, según diferentes edades, sexo, estado conyugal, mortalidad y distribución geográfica, constituye un elemento importante para la generación y organización de los datos demográficos y sociales.

Esta información es utilizada para diferentes fines: desde las personas que tienen la inquietud por conocer su desarrollo, hasta aquellas instituciones públicas y privadas que tienen a su cargo formular políticas, planes y programas de trabajo encaminados a resolver los problemas de la población.

Dentro de este marco de necesidades, se encuentra el interés por la generación y divulgación de las estadísticas vitales, con el propósito de disponer de datos relativos a sus categorías fundamentales y estar en posibilidad de analizar el comportamiento y tendencias del crecimiento de la población.

México no se cuenta con un registro nacional de cáncer, es decir, hoy en día no se sabe el número de personas que con cáncer en la Republica Mexicana. Sin

embargo se cuentan con otros registros que son de utilidad para poder dar una estimación primero de los casos incidentes y después de los casos prevalentes.

En México se cuenta con un excelente registro de las estadísticas vitales. Las estadísticas vitales captadas por el INEGI, mediante el aprovechamiento de los registros administrativos de diversas instituciones públicas, son el resultado del recuento de los hechos ocurridos en la vida de la población, como son: nacimientos, matrimonios, divorcios, defunciones y muertes fetales.

Las estadísticas vitales son elementos básicos para el análisis demográfico de la situación de un país, así como uno de los requisitos para poder llevar a cabo la planificación del desarrollo económico y social. Ya que proporcionan información sobre la tendencia del crecimiento natural de la población basándose en las tasas de natalidad y mortalidad; sobre la conducta de sus componentes, su distribución geográfica y mediante su agregación a lo largo del tiempo, sobre el tamaño de la población y su estructura. Por otro lado, permite identificar a los grupos demandantes de servicios médicos, educación, vivienda, etc.

Las estadísticas vitales se componen de:

- **Estadísticas de nacimiento.** La intención de estudiar las características que identifican al nacimiento es conocer la frecuencia con que ocurren estos hechos en el país y permitir medir su severidad; así como las condiciones sociales y económicas en que se desarrolla este hecho, debido a que una vez obtenido su volumen y desglose, es posible conocer, entre otros aspectos, la efectividad de los programas de salud materno-infantil, de planificación familiar, de las campañas de registro, así como detectar las necesidades de servicios y recursos médicos.

- **Estadísticas de defunciones.** El propósito de la estadística de defunciones es producir en forma continua, información que permita conocer y comparar el volúmen, tendencias y características de la mortalidad en los diferentes ámbitos geográficos del país, lo que constituye un insumo para el análisis y evaluación de acciones dirigidas a la elaboración de programas de salud pública, para controlar enfermedades infecciosas y epidemiológicas, prevención de accidentes y en el estudio de diferencias de la mortalidad por edad, sexo y causa básica de la defunción.

- **Estadísticas de muertes fetales.** El objetivo de la estadística de Muertes Fetales es generar información que permita conocer la frecuencia con que ocurren estas muertes; la situación social y económica de los padres, sobre el estado de salud de la madre y del producto, así como las causas que originan la muerte fetal

- **Estadísticas de matrimonios.** Con la información que se genera de la estadística de matrimonios es posible conocer el volumen de los matrimonios civiles registrados en el país, además de las características demográficas y socioeconómicas de la población involucrada; esto permite conocer

en qué medida se da la formación de nuevas familias y a qué edad llegan las parejas al matrimonio, aspectos que están relacionados con el análisis de la fecundidad. Asímismo esta información permite proyectar la demanda de necesidades básicas como: alimentación, vivienda, servicios de salud y planificación familiar, entre otros aspectos.

- **Estadísticas divorcios.** La estadística de divorcios tiene como finalidad presentar la frecuencia con que ocurren las disoluciones legales de los matrimonios registrados en el país, lo que permite conocer las causas y factores fundamentales que influyen para que ocurra este hecho; esta información proporciona elementos que contribuyen a definir, elaborar y aplicar programas asistenciales de apoyo a menores de edad, hijos de padres divorciados o separados, programas de orientación y fomento al ejercicio de la paternidad responsable.

Por otro lado, en México también se cuentan con proyecciones de población. La Secretaría General del Consejo Nacional de Población (CONAPO), tiene entre sus responsabilidades el analizar, evaluar, sistematizar y producir información sobre los fenómenos demográficos, así como elaborar proyecciones de población.

Actualmente se cuentan con las proyecciones de población para el horizonte 2010-2030. Dicha información es necesaria y relevante para llevar a cabo la planeación demográfica, económica y social del país, al mismo tiempo que una herramienta de conocimiento valiosa para estimar múltiples requerimientos futuros en servicios e infraestructura así como otras necesidades sociales.

Cabe mencionar que las proyecciones de población se actualizan a partir de la disponibilidad de un nuevo censo de población y vivienda, o un conteo de población.

3.1.1. Estimación de casos incidentes.

Hasta el momento se cuenta con información de las defunciones generales que a su vez están divididas por causa de la muerte, edad y sexo. Así mismo, se tienen las proyecciones de población hasta el 2030.

Si suponemos que se cuenta con una tasa de incidencia en cáncer, entonces el número de casos incidentes de cáncer en México al año 2020 se puede obtener como:

$$_{edad}CEI_t^j = \frac{(_{edad}P_t^j)(_{edad}TI_t^j)}{m}$$

donde:

CEI_t son los casos esperados incidentes de cáncer al tiempo t.
P_t es la población en riesgo al tiempo t.
TI_t es la tasa de incidencia a tiempo t.
m es la proporción de la tasa.
edad es la edad del individuo.
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

Desafortunadamente no es posible calcular o estimar la tasa de incidencia con datos nacionales.

Por lo anterior se utilizaron los datos de *The GLOBOCAN project*, la cual proporciona tasas de incidencia en cáncer para los años 2002, 2008 y 2012.

Las tasas reportadas por *The GLOBOCAN project* están divididas por grupo de edad y sexo; los grupos de edad son: $[0-14]$, $[15-39]$, $[40-44]$, $[45-49]$, $[50-54]$, $[55-59]$, $[60-64]$, $[65-69]$, $[70-74]$ y $[75-75+]$. Además, todas las tasas son por cada $100,000$ habitantes ($m = 100,000$).

Entonces, definimos x como:

$$x = \begin{cases} 1 & \text{si } edad \in [0-14] \\ 2 & \text{si } edad \in [15-39] \\ 3 & \text{si } edad \in [40-44] \\ 4 & \text{si } edad \in [45-49] \\ 5 & \text{si } edad \in [50-54] \\ 6 & \text{si } edad \in [55-59] \\ 7 & \text{si } edad \in [60-64] \\ 8 & \text{si } edad \in [65-69] \\ 9 & \text{si } edad \in [70-74] \\ 10 & \text{si } edad \in [75-75+] \end{cases}$$

Sin embargo, solo se cuenta con tasas para los años 2002, 2008 y 2012 y para algunos grupos de edad se notó un crecimiento lineal a través del tiempo. Este último punto no tiene sentido práctico, ya que como se mencionó en el capítulo 1, el cáncer no es una enfermedad infecciosa o viral, entonces pensar en un crecimiento lineal o exponencial sobreestimaria las tasas de incidencia, así que se decidió suponer un comportamiento logarítmico para las tasas de incidencia a lo largo del tiempo y ajustar una curva mediante el método de **mínimos cuadrados ordinarios**. Es decir, suponemos que:

$$TI_t = \alpha_0 + \alpha_1 ln(t), \quad \forall\ t \in \{2002, 2008, 2012\}$$

Por tanto buscamos $\hat{\alpha_0}$ y $\hat{\alpha_1}$ tales que se minimice la siguiente función:

$$S(\alpha_0, \alpha_1)|_{\hat{\alpha_0}, \hat{\alpha_1}} = \sum_{i=1}^{n} (TI_{t_i} - \hat{\alpha_0} + \hat{\alpha_1} ln(t_i))^2$$

donde n es el número de datos que se tienen.

Entonces, los estimadores por mínimos cuadrados deben de cumplir que:

$$\left.\frac{\partial F}{\partial \alpha_0}\right|_{\hat{\alpha_0}, \hat{\alpha_1}} = -\sum_{i=1}^{n} TI_{t_i} + n\hat{\alpha_0} + \hat{\alpha_1} \sum_{i=1}^{n} ln(t_i) = 0$$

y

$$\left.\frac{\partial F}{\partial \alpha_1}\right|_{\hat{\alpha_0}, \hat{\alpha_1}} = -\sum_{i=1}^{n} TI_{t_i} ln(t_i) + \hat{\alpha_0} \sum_{i=1}^{n} ln(t_i) + \hat{\alpha_1} \sum_{i=1}^{n} (ln(t_i))^2 = 0$$

Simplificando el sistema ecuaciones:

$$n\hat{\alpha_0} + \hat{\alpha_1} \sum_{i=1}^{n} ln(t_i) = \sum_{i=1}^{n} TI_{t_i}$$

$$\hat{\alpha_0} \sum_{i=1}^{n} ln(t_i) + \hat{\alpha_1} \sum_{i=1}^{n} (ln(t_i))^2 = \sum_{i=1}^{n} TI_{t_i} ln(t_i)$$

El cual es un sistema lineal no singular de 2x2. Por lo tanto, la solución está dada por:

$$\hat{\alpha_0} = \frac{\sum_{i=1}^{n} TI_{t_i} - \hat{\alpha_1} \sum_{i=1}^{n} ln(t_i)}{n}$$

y

$$\hat{\alpha_1} = \frac{n \sum_{i=1}^{n} TI_{t_i} ln(t_i) - \sum_{i=1}^{n} TI_{t_i} \sum_{i=1}^{n} ln(t_i)}{n \sum_{i=1}^{n} (ln(t_i))^2 - \left(\sum_{i=1}^{n} ln(t_i)\right)^2}$$

De esta forma se obtiene la curva logarítmica más cercana a los datos por el método de mínimos cuadrados ordínarios. De esta forma se puede conocer $TI_t \; \forall \; t \in \{2002, ..., 2020\}$ [**Apéndice B**].

Los valores de $\hat{\alpha_0}$ y $\hat{\alpha_1}$ para cada ajuste fueron:

Sexo	Grupo de edad	$\hat{\alpha_0}$	$\hat{\alpha_0}$
Hombre	$[0-14]$	13.5648	-0.9426
Hombre	$[15-39]$	33.1802	11.2410
Hombre	$[40-44]$	72.5339	23.7125
Hombre	$[45-49]$	98.5348	34.5124
Hombre	$[50-54]$	174.0685	64.3023
Hombre	$[55-59]$	292.8160	108.3696
Hombre	$[60-64]$	454.7886	157.6150
Hombre	$[65-69]$	642.4992	208.4748
Hombre	$[70-74]$	916.7284	290.8906
Hombre	$[75-75+]$	1,231.9780	403.1692
Mujer	$[0-14]$	10.9261	3.5934
Mujer	$[15-39]$	48.9699	14.5681
Mujer	$[40-44]$	162.8379	49.3982
Mujer	$[45-49]$	225.3899	68.2002
Mujer	$[50-54]$	270.9168	78.2363
Mujer	$[55-59]$	336.5271	99.5321
Mujer	$[60-64]$	436.5433	135.6110
Mujer	$[65-69]$	525.2461	167.2018
Mujer	$[70-74]$	634.2354	204.8119
Mujer	$[75-75+]$	712.9930	226.1347

Si suponemos que toda la población es propensa a desarrollar cáncer, entonces el número esperado de casos incidentes se calcula como:

$$ {}_xCEI_t^j = \frac{({}_xP_t^j)({}_xTI_t^j)}{100,000} $$

donde:

CEI_t son los casos esperados incidentes de cáncer al tiempo t.
P_t es la población estimada por CONAPO a tiempo t.
TI_t es la tasa de incidencia a tiempo t.
$100,000$ es la proporción de la tasa.
x es el grupo de edad ($x \in \{1, 2, ..., 10\}$).
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

3.1.2. Estimación de casos prevalentes.

Para estimar el número de casos prevalentes para el grupo de edad x y el sexo j a tiempo t se calcula:

$$ {}_xCEP_t^j = {}_xCEI_t^j - {}_xDc_t^j + {}_xCEP_{t-1}^j $$

donde:

CEP_t son los casos esperados prevalentes de cáncer al tiempo t.
CEI_t son los casos esperados incidentes de cáncer al tiempo t.
Dc_t son las defunciones causadas por cáncer al tiempo t.
x es el grupo de edad ($x \in \{1, 2, ..., 10\}$)
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer)

Para obtener ${}_xDc_t^j \ \forall \ t \in \{2002, ..., 2020\}$ se utilizarón métodos de proyección y pronósticos estadísticos mediante el siguiente criterio: Primero se ajustan modelos de regresión lineal simple. Si el modelo resulta se adecuado, i.e., todos los parámetros son significativos, no se viola supuesto alguno y $R^2 > 80\,\%$, entonces con ese modelo se hacen las predicciones. En caso contrario, los pronósticos se harán ajustando un modelo de series de tiempo. El mismo criterio se utilizó para todos los ajustes posteriores.

Dado el criterio anterior, estos fueron los modelos ocupados para proyectar las defunciones:

Sexo	Grupo de edad	Modelo	Parámetros
Hombre	$[0-14]$	IMA$(2,1)$	$\hat{\theta_1} = -0{,}989$
Hombre	$[15-39]$	Regresión	$\hat{\beta_0} = -87{,}070.6$; $\hat{\beta_1} = 44.7$
Hombre	$[40-44]$	Regresión	$\hat{\beta_0} = -47{,}006.5$; $\hat{\beta_1} = 23.8$
Hombre	$[45-49]$	Regresión	$\hat{\beta_0} = -65{,}070.5$; $\hat{\beta_1} = 33.1$
Hombre	$[50-54]$	Regresión	$\hat{\beta_0} = -116{,}426.5$; $\hat{\beta_1} = 58.9$
Hombre	$[55-59]$	Regresión	$\hat{\beta_0} = -152{,}739.8$; $\hat{\beta_1} = 77.3$
Hombre	$[60-64]$	Regresión	$\hat{\beta_0} = -159{,}694.1$; $\hat{\beta_1} = 81.1$
Hombre	$[65-69]$	Regresión	$\hat{\beta_0} = -143{,}847.3$; $\hat{\beta_1} = 73.6$
Hombre	$[70-74]$	Regresión	$\hat{\beta_0} = -183{,}932.9$; $\hat{\beta_1} = 93.9$
Hombre	$[75-75+]$	Regresión	$\hat{\beta_0} = -749{,}952.1$; $\hat{\beta_1} = 379.4$

Sexo	Grupo de edad	Modelo	Parámetros
Mujer	$[0-14]$	IMA$(2,1)$	$\hat{\theta_1} = -0.967$
Mujer	$[15-39]$	Regresión	$\hat{\beta_0} = -57{,}745.6$; $\hat{\beta_1}$ =30.2
Mujer	$[40-44]$	Regresión	$\hat{\beta_0} = -37{,}766.8$; $\hat{\beta_1}$ =16.6
Mujer	$[45-49]$	Regresión	$\hat{\beta_0} = -102{,}078.8$; $\hat{\beta_1}$ =52.1
Mujer	$[50-54]$	Regresión	$\hat{\beta_0} = -168{,}430.9$; $\hat{\beta_1}$ =85.3
Mujer	$[55-59]$	Regresión	$\hat{\beta_0} = -205{,}755.7$; $\hat{\beta_1}$ =104.1
Mujer	$[60-64]$	Regresión	$\hat{\beta_0} = -179{,}480.6$; $\hat{\beta_1}$ =91.2
Mujer	$[65-69]$	Regresión	$\hat{\beta_0} = -148{,}144.8$; $\hat{\beta_1}$ =75.7
Mujer	$[70-74]$	Regresión	$\hat{\beta_0} = -156{,}360.1$; $\hat{\beta_1}$ =79.8
Mujer	$[75-75+]$	Regresión	$\hat{\beta_0} = -565{,}238.1$; $\hat{\beta_1}$ =286.4

3.2. Estimación del costo directo.

El modelo que se utilizó para estimar el costo directo es conocido como ***Modelo Costo-Enfermedad***, el cual define el costo directo como el producto de los casos esperados incidentes a tiempo t por el costo total de antención de la enfermedad. En otras palabras

$$ {}_xCD_t^j = ({}_xCEI_t^j)(CT) $$

donde:

CD_t es el costo directo por cáncer a tiempo t.
CEI_t son los casos esperados incidentes de cáncer a tiempo t.
CT es el costo total de atención por cáncer.
x es el grupo de edad ($x \in \{1, 2, ..., 10\}$)
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

En este caso CT no se dividió por edad o sexo porque se estimó un valor único del costo de atención por cáncer ya que el costo de atención depende tanto de la edad, sexo y tipo de cáncer así como de la etapa clínica. Por lo anterior, el costo de atención se estimó como el costo promedio de los cánceres cubiertos por el Seguro Popular, los cuales son: cáncer de esófago, cáncer de mama, cáncer de páncreas, cáncer de cuello uterino, cáncer de colon, cáncer de recto, cáncer de próstata, cáncer de pulmón y cáncer de estómago, los cuales representan cerca del 70 % de las muertes por cáncer en México.

3.3. Estimación del costo indirecto.

El modelo que se utilizó para estimar el costo indirecto se conoce como ***Modelo de Capital Humano***. Dicho modelo es utilizado tanto en economía como en sociología y postula que a partir de la información a la que tienen acceso los agentes económicos, deciden sobre sus pautas de inversión, con el fin aumentar su

productividad y obtener unos mayores ingresos en el futuro. Sin embargo, la inversión no tienen por qué realizarse exclusivamente sobre bienes materiales, también puede producirse sobre la propia capacidad de la persona. El punto clave reside en que esa inversión, esa renuncia a un consumo presente, se traduzca en un aumento de productividad futuro y esto, tendrá su reflejo en un aumento futuro de su riqueza. Por lo tanto, ***el costo indirecto será aquel costo generado por la falta de productividad de la persona a causa de la enfermedad***.

El costo indirecto se divide en:

- Ingreso perdido por muerte prematura.
- Subsidio por incapacidad temporal.
- Pensión por invalidez.
- Costo de oportunidad del cuidador.

3.3.1. Ingreso perdido por muerte prematura.

Nota: en esta sección, los grupos de edad estarán determinados por quinquenios a partir de 15 y hasta 64 años. La notación para dicho cambio en los grupos de edad será x^* donde $x^* \in \{1^*, ..., 10^*\}$ de tal forma que i^* hace referencia al i-ésimo quinquenio a partir de 15 años. Este cambio sólo se puede hacer en este apartado, ya que para el cálculo del ingreso perdido por muerte prematura no se necesitan el cálculo de las casos esperados en cáncer, en consecuencia, tampoco los grupos de edad dictados por
The GLOBOCAN project.

El **ingreso perdido por muerte prematura**, se define como el ingreso que dejará de percibir la persona a causa de la muerte por la enfermedad en cuestión. Se calcula como el valor presente actuarial del ingreso anual por el número de defunciones al año t. Es decir:

$$_{x^*}IPMP_t^j = ([_{x^*}\,\%PEA_t^j][_{x^*}\,\%PEAO][_{x^*}Dc_t^j])(_{x^*}IA_t^j)\ddot{a}^j_{\bar{x}^*:\overline{65-\bar{x}^*}\rceil}$$

donde:

$IPMP_t$ es el ingreso perdido por muerte prematura en cáncer a tiempo t.
$\%PEA$ es el porcentaje de la población económicamente activa a tiempo t.
$\%PEAO$ es el porcentaje de la población económicamente activa que está ocupada a tiempo t.
Dc_t son las defunciones por cáncer al tiempo t.
IA_t es el ingreso anual a tiempo t.
x^* es el grupo de edad $x^* \in \{1^*, ..., 10^*\}$.
$\bar{x}^*$ es la edad promedio del grupo de edad x^*.
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

Nótese que con el primer paréntesis $([_{x^*}\,\%PEA_t^j][_{x^*}\,\%PEAO][_{x^*}Dc_t^j])$ se está calculando *el número esperado de muertes por cáncer dentro de la población económicamente activa y ocupada a tiempo t.*

Los modelos usados para pronosticar la $\%PEA$ al año 2020 fueron:

Sexo	Grupo de edad	Modelo	Parámetros
Hombre	$[15-19]$	$AR(5)$	$\hat{\phi}_1 = -0{,}391$ $\hat{\phi}_3 = -0{,}241$ $\hat{\phi}_4 = -0{,}400$ $\hat{\phi}_5 = -0{,}569$
Hombre	$[20-24]$	$ARMA(3,2)$	$\hat{\phi}_1 = 0{,}261$ $\hat{\phi}_2 = 0{,}424$ $\hat{\phi}_3 = -0{,}837$ $\hat{\theta}_2 = -0{,}977$
Hombre	$[25-29]$	$AR(5)$	$\hat{\phi}_2 = -0{,}190$ $\hat{\phi}_3 = -0{,}248$ $\hat{\phi}_4 = -0{,}306$ $\hat{\phi}_5 = -0{,}537$
Hombre	$[30-34]$	$AR(5)$	$\hat{\phi}_1 = -0{,}480$ $\hat{\phi}_4 = -0{,}915$ $\hat{\phi}_5 = -0{,}488$
Hombre	$[35-39]$	$AR(4)$	$\hat{\phi}_4 = -0{,}770$
Hombre	$[40-44]$	$AR(6)$	$\hat{\phi}_2 = -0{,}837$ $\hat{\phi}_4 = -0{,}924$ $\hat{\phi}_6 = -0{,}851$
Hombre	$[45-49]$	$ARMA(5,2)$	$\hat{\phi}_1 = -0{,}537$ $\hat{\phi}_3 = -0{,}294$ $\hat{\phi}_4 = -0{,}648$ $\hat{\phi}_5 = -0{,}725$ $\hat{\theta}_2 = -0{,}967$
Hombre	$[50-54]$	$AR(6)$	$\hat{\phi}_1 = -0{,}196$ $\hat{\phi}_2 = -0{,}350$ $\hat{\phi}_3 = -0{,}891$ $\hat{\phi}_5 = -0{,}711$ $\hat{\phi}_6 = -0{,}577$
Hombre	$[55-59]$	$ARMA(3,2)$	$\hat{\phi}_1 = 0{,}393$ $\hat{\phi}_2 = 0{,}438$ $\hat{\phi}_3 = -0{,}954$ $\hat{\theta}_2 = -0{,}978$
Hombre	$[60-64]$	$ARMA(3,2)$	$\hat{\phi}_1 = 0{,}311$ $\hat{\phi}_2 = 0{,}465$ $\hat{\phi}_3 = -0{,}846$ $\hat{\theta}_2 = -0{,}979$

Sexo	Grupo de edad	Modelo	Parámetros
Mujer	$[15-19]$	$ARMA(5,2)$	$\hat{\phi_1} = -0{,}914$ $\hat{\phi_4} = -0{,}945$ $\hat{\phi_5} = -0{,}920$ $\hat{\theta_2} = -0{,}865$
Mujer	$[20-24]$	$ARMA(3)$	$\hat{\phi_1} = -1{,}018$ $\hat{\phi_2} = -0{,}937$ $\hat{\phi_3} = -0{,}872$
Mujer	$[25-29]$	$AR(6)$	$\hat{\phi_1} = -0{,}497$ $\hat{\phi_3} = -0{,}503$ $\hat{\phi_4} = -0{,}559$ $\hat{\phi_5} = -0{,}572$ $\hat{\phi_6} = -0{,}572$
Mujer	$[30-34]$	$ARMA(5,2)$	$\hat{\phi_1} = -0{,}836$ $\hat{\phi_4} = -0{,}515$ $\hat{\phi_4} = -0{,}677$ $\hat{\theta_2} = -0{,}958$
Mujer	$[35-39]$	$ARMA(5,1)$	$\hat{\phi_1} = 0{,}146$ $\hat{\phi_3} = -0{,}549$ $\hat{\phi_5} = -0{,}452$ $\hat{\theta_1} = -0{,}934$
Mujer	$[40-44]$	$ARMA(5,2)$	$\hat{\phi_1} = -0{,}400$ $\hat{\phi_3} = -0{,}414$ $\hat{\phi_4} = -0{,}446$ $\hat{\phi_5} = -0{,}656$ $\hat{\theta_2} = -0{,}971$
Mujer	$[45-49]$	$ARMA(5,2)$	$\hat{\phi_3} = -0{,}529$ $\hat{\phi_5} = -0{,}468$ $\hat{\theta_2} = -1{,}093$
Mujer	$[50-54]$	$ARMA(5,2)$	$\hat{\phi_1} = -0{,}579$ $\hat{\phi_3} = -0{,}365$ $\hat{\phi_4} = -0{,}580$ $\hat{\phi_5} = -0{,}634$ $\hat{\theta_2} = -0{,}948$
Mujer	$[55-59]$	$ARMA(5,1)$	$\hat{\phi_2} = -0{,}402$ $\hat{\phi_3} = -0{,}458$ $\hat{\phi_5} = -0{,}905$ $\hat{\theta_1} = -1{,}001$
Mujer	$[60-64]$	$ARMA(5,1)$	$\hat{\phi_3} = -0{,}564$ $\hat{\phi_4} = -0{,}240$ $\hat{\phi_5} = -0{,}426$ $\hat{\theta_1} = -0{,}943$

Por otro lado, los modelos usados para pronosticar la $\%PEAO$ al año 2020 fueron:

Sexo	Grupo de edad	Modelo	Parámetros
Hombre	$[15-19]$	$ARMA(4,1)$	$\hat{\phi}_4 = -0{,}997$ $\hat{\theta}_1 = -0{,}966$
Hombre	$[20-24]$	$ARMA(4,1)$	$\hat{\phi}_4 = -0{,}980$ $\hat{\theta}_1 = -0{,}943$
Hombre	$[25-29]$	$ARMA(4,1)$	$\hat{\phi}_4 = -0{,}971$ $\hat{\theta}_1 = -0{,}982$
Hombre	$[30-34]$	$AR(4)$	$\hat{\phi}_4 = -0{,}938$
Hombre	$[35-39]$	$ARMA(4,1)$	$\hat{\phi}_4 = -0{,}987$ $\hat{\theta}_1 = -0{,}997$
Hombre	$[40-44]$	$ARMA(5,2)$	$\hat{\phi}_1 = -0{,}682$ $\hat{\phi}_3 = -0{,}203$ $\hat{\phi}_4 = -0{,}732$ $\hat{\phi}_5 = -0{,}846$ $\hat{\theta}_2 = -0{,}931$
Hombre	$[45-49]$	$ARMA(5,1)$	$\hat{\phi}_2 = -0{,}335$ $\hat{\phi}_4 = -0{,}694$ $\hat{\phi}_5 = -0{,}427$ $\hat{\theta}_1 = -0{,}976$
Hombre	$[50-54]$	$ARMA(6,1)$	$\hat{\phi}_1 = -0{,}338$ $\hat{\phi}_2 = -0{,}446$ $\hat{\phi}_3 = -0{,}506$ $\hat{\phi}_4 = -0{,}487$ $\hat{\phi}_5 = -0{,}530$ $\hat{\phi}_6 = -0{,}674$ $\hat{\theta}_1 = -0{,}991$
Hombre	$[55-59]$	$AR(5)$	$\hat{\phi}_4 = -0{,}869$ $\hat{\phi}_5 = -0{,}140$
Hombre	$[60-64]$	$ARMA(5,1)$	$\hat{\phi}_3 = -0{,}332$ $\hat{\phi}_4 = 0{,}477$ $\hat{\phi}_5 = -0{,}377$ $\hat{\theta}_1 = -0{,}959$

Sexo	Grupo de edad	Modelo	Parámetros
Mujer	$[15-19]$	$ARMA(5,1)$	$\hat{\phi_1} = -0{,}417$ $\hat{\phi_3} = -0{,}109$ $\hat{\phi_4} = -0{,}855$ $\hat{\phi_5} = -0{,}512$ $\hat{\theta_1} = -0{,}986$
Mujer	$[20-24]$	$ARMA(4)$	$\hat{\phi_4} = -0{,}908$
Mujer	$[25-29]$	$AR(4,1)$	$\hat{\phi_3} = -0{,}223$ $\hat{\phi_4} = -0{,}811$ $\hat{\theta_1} = -0{,}978$
Mujer	$[30-34]$	$ARMA(4,1)$	$\hat{\phi_4} = -0{,}982$ $\hat{\theta_1} = -0{,}981$
Mujer	$[35-39]$	$ARMA(5,1)$	$\hat{\phi_1} = 0{,}380$ $\hat{\phi_2} = -0{,}584$ $\hat{\phi_3} = -0{,}542$ $\hat{\phi_4} = 0{,}302$ $\hat{\phi_5} = -0{,}933$ $\hat{\theta_1} = -0{,}984$
Mujer	$[40-44]$	$ARMA(5,1)$	$\hat{\phi_1} = 0{,}114$ $\hat{\phi_2} = -0{,}266$ $\hat{\phi_3} = -0{,}526$ $\hat{\phi_5} = -0{,}772$ $\hat{\theta_1} = -0{,}988$
Mujer	$[45-49]$	$ARMA(6,1)$	$\hat{\phi_2} = -0{,}958$ $\hat{\phi_4} = -0{,}986$ $\hat{\phi_6} = -0{,}954$ $\hat{\theta_2} = -1{,}000$
Mujer	$[50-54]$	$AR(5)$	$\hat{\phi_1} = -0{,}950$ $\hat{\phi_4} = -0{,}970$ $\hat{\phi_5} = -0{,}962$
Mujer	$[55-59]$	$AR(5)$	$\hat{\phi_1} = -0{,}460$ $\hat{\phi_3} = -0{,}273$ $\hat{\phi_4} = -0{,}638$ $\hat{\phi_5} = -0{,}660$
Mujer	$[60-64]$	$AR(6)$	$\hat{\phi_2} = -0{,}673$ $\hat{\phi_4} = -0{,}850$ $\hat{\phi_6} = -0{,}660$

IA_t se calculó de dos formas para interpretar dos escenarios:

- Utilizando el Salario Base de Cotizaciones del IMSS para el Seguro de Enfermedades y Maternidad (SEM) y para el Seguro de Invalidez y Vida (SIV), ambos deflactados a pesos corrientes de diciembre del 2013 y proyectado al 2020.
- Utilizando el ingreso promedio reportado por la Encuesta Nacional de Ocupación y Empleo (durante el segundo trimestre de cada año), deflactado a pesos corrientes de diciembre del 2013 y proyectado al 2020.

De esta forma, al final se tendrán dos cantidades, dentro de las cuales estará el ingreso perdido por muerte prematura ya que en la ENOE se registra el salario autorreportado de las personas, y que en general tienden a mentir. Por otro lado el SBC del SEM (y del SIV) es un cálculo propio del IMSS basado en el reporte de las empresas, pero se deja fuera a todos aquellos no afiliados al IMSS.

Los modelos usados para pronosticar al SEM y SIV fueron:

SBC	Modelo	Parámetros
SEM	$ARI(3,2)$	$\hat{\phi_3} = -0{,}6942$
SIV	$ARI(3,2)$	$\hat{\phi_3} = -0{,}5937$

Cabe destacar, que para los datos reportados por la ENOE se tienen datos a partir del 2005, así que para los años anteriores se supuso un crecimiento salarial arítmetico del $\$1,773.00$, el cual es el promedio del crecimiento salarial de hombres y mujeres. Después de considerar dichos valores, se hicieron los respectivos pronósticos al año 2020. Lo anterior se hizo para evitar errores contextuales, ya que si se utilizaba el método usual, las proyecciones indicaban que en algún punto del tiempo anterior al 2005, las mujeres tenían un ingreso anual más alto que el de los hombres, lo cual sabemos que históricamente es falso.

Los modelos usados para proyectar el ingreso promedio anual de la ENOE fueron:

Sexo	Grupo de edad	Modelo	Parámetros
Hombre	$[15-19]$	Regresión	$\hat{\beta}_0 = -3{,}692{,}681.6$; $\hat{\beta}_1 = 1{,}852.7$
Hombre	$[20-24]$	Regresión	$\hat{\beta}_0 = -4{,}643{,}874.2$; $\hat{\beta}_1 = 2{,}332.7$
Hombre	$[25-29]$	Regresión	$\hat{\beta}_0 = -5{,}403{,}953.6$; $\hat{\beta}_1 = 2{,}716.2$
Hombre	$[30-34]$	Regresión	$\hat{\beta}_0 = -5{,}351{,}202.3$; $\hat{\beta}_1 = 2{,}692.6$
Hombre	$[35-39]$	Regresión	$\hat{\beta}_0 = -5{,}271{,}530.6$; $\hat{\beta}_1 = 2{,}654.2$
Hombre	$[40-44]$	Regresión	$\hat{\beta}_0 = -5{,}580{,}378.6$; $\hat{\beta}_1 = 2{,}809.6$
Hombre	$[45-49]$	Regresión	$\hat{\beta}_0 = -5{,}556{,}030.3$; $\hat{\beta}_1 = 2{,}797.9$
Hombre	$[50-54]$	Regresión	$\hat{\beta}_0 = -5{,}499{,}489.1$; $\hat{\beta}_1 = 2{,}769.4$
Hombre	$[55-59]$	Regresión	$\hat{\beta}_0 = -6{,}316{,}074.9$; $\hat{\beta}_1 = 3{,}172.9$
Hombre	$[60-64]$	Regresión	$\hat{\beta}_0 = -4{,}911{,}727.1$; $\hat{\beta}_1 = 2{,}469.5$

Sexo	Grupo de edad	Modelo	Parámetros
Mujer	$[14-19]$	Regresión	$\hat{\beta}_0 = -3{,}419{,}439.5$; $\hat{\beta}_1 = 1{,}714.4$
Mujer	$[20-24]$	Regresión	$\hat{\beta}_0 = -4{,}167{,}876.3$; $\hat{\beta}_1 = 2{,}092.2$
Mujer	$[25-29]$	Regresión	$\hat{\beta}_0 = -4{,}765{,}714.1$; $\hat{\beta}_1 = 2{,}394.2$
Mujer	$[30-34]$	Regresión	$\hat{\beta}_0 = -4{,}667{,}065.3$; $\hat{\beta}_1 = 2{,}345.5$
Mujer	$[35-39]$	Regresión	$\hat{\beta}_0 = -4{,}419{,}970.1$; $\hat{\beta}_1 = 2{,}222.2$
Mujer	$[40-44]$	Regresión	$\hat{\beta}_0 = -4{,}388{,}556.3$; $\hat{\beta}_1 = 2{,}207.4$
Mujer	$[45-49]$	Regresión	$\hat{\beta}_0 = -4{,}443{,}321.5$; $\hat{\beta}_1 = 2{,}235.0$
Mujer	$[50-54]$	Regresión	$\hat{\beta}_0 = -4{,}775{,}629.9$; $\hat{\beta}_1 = 2{,}399.4$
Mujer	$[55-59]$	Regresión	$\hat{\beta}_0 = -4{,}607{,}218.7$; $\hat{\beta}_1 = 2{,}312.1$
Mujer	$[60-64]$	Regresión	$\hat{\beta}_0 = -4{,}775{,}818.4$; $\hat{\beta}_1 = 2{,}393.3$

3.3.2. Subsidio por incapacidad temporal.

Nota. Cuando se haga referencia a *asalariados* o *personas asalariadas*, se pensará exclusivamente en la población *afiliada al IMSS*.

El subsidio por incapacidad temporal se define como el subsidio que dará el Instituto Mexicano del Seguro Social por la incapacidad del trabajador para laborar debido a enfermedad.

La Ley del Seguro Social, en su artículo 96 dice que: *en caso de enfermedad no profesional, el asegurado tendrá derecho a un subsidio en dinero que se otorgará cuando la enfermedad lo incapacite para el trabajo. El subsidio se pagará a partir del cuarto día del inicio de la incapacidad, mientras dure esta y hasta por el término de cincuenta y dos semanas.*

Por otro lado, el artículo 98 indica que: *el subsidio en dinero que se otorgue a los asegurados será igual al* ***sesenta por ciento del último salario diario de cotización.***

Por tanto, el subsidio por incapacidad laboral se calculará como:

$$_xSIT_t^j = ([_xCEI_t^j][\,\%As_t^j])(270-4)(60\,\%SEM_t^j)$$

donde:

SIT_t es el subsidio por incapacidad temporal a tiempo t .
CEI_t son los casos esperados incidentes de cáncer al tiempo t.
$\%As_t$ es el porcentaje de la población asalariada a tiempo t.
$270-4$ es el número promedio de días de incapacidad por cáncer menos los 4 días que marca la ley.
SEM_t es el salario base de cotizaciones correspondiente al seguro de enfermedades y maternidad a tiempo t.
x es el grupo de edad ($x \in \{2, ..., 7\}$)
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer)

Nótese que con el primer paréntesis ($[_xCEI_t^j][\,\%As_t]$) se está calculando *el número esperado de personas asalariadas con cáncer al año t.*

En este caso el enfoque fue hacia la población asalariada ya que este subsidio es por parte del IMSS.

El porcentaje de la población asalariada a tiempo t ($\%As_t$) se obtuvo por medio de los datos reportados por la ENOE (a través de la tasa de ocupación asalariada). Después se hizo la proyección al año 2020 utilizando los siguientes modelos:

Sexo	Modelo	Parámetros
Hombre	$ARMA(2,1)$	$\hat{\phi_1}$ =1.021 $\hat{\phi_2} = -0.770$ $\hat{\theta_1} = -0.989$
Mujer	$ARMA(4,1)$	$\hat{\phi_1} = -0.414$ $\hat{\phi_2} = -0.605$ $\hat{\phi_3} = -0.540$ $\hat{\phi_4} = -0.652$ $\hat{\theta_1} = -0.959$

3.3.3. Pensión por invalidez.

La pensión por invalidez se define como la pensión que otorgará el Instituto Mexicano de Seguro Social por invalidez del trabajador a causa de accidente laboral o enfermedad.

La Ley del Seguro Social, en su artículo 119 estipula: *para los efectos de esta ley existe invalidez cuando el asegurado se halle imposibilitado para procurarse, mediante un trabajo igual, una remuneración superior al cincuenta por ciento de su remuneración habitual percibida durante el último año de trabajo y que esa imposibilidad derive de una enfermedad o accidente no profesionales. La declaración de invalidez debera ser realizada por el instituto mexicano del seguro social.*

Por otro lado, el artículo 124 dicta que: *los asegurados que soliciten el otorgamiento de una pensión de invalidez y los inválidos que se encuentren disfrutándola, deberán sujetarse a las investigaciones de carácter médico, social y económico que el instituto estime necesarias, para comprobar si existe o subsiste el estado de invalidez.*

Además, el artículo 141 ordena que: *la cuantía de la pensión por invalidez será igual a una cuantía básica del treinta y cinco por ciento del promedio de los salarios correspondientes a las ultimas quinientas semanas de cotizacion anteriores al otorgamiento de la misma, actualizadas conforme al Indice Nacional de Precios al Consumidor, mas las asignaciones familiares y ayudas asistenciales.*

En consecuencia, la pensión por invalidez se calculará como:

$$ {}_xPI_t^j = (35\,\%\bar{SIV}_t)([{}_xCEI_t^j][\,\%As_t^j])(1 - {}_xf^j)\ \ddot{a}_{\bar{x}}^j $$

donde:

PI_t es la pensión por invalidez a tiempo t.
$\bar{SIV}_t$ es el salario base de cotizaciones promedio de las últimas 500 semanas correspondiente al seguro de invalidez y vida a tiempo t.
CEI_t son los casos esperados incidentes de cáncer al tiempo t.
$\%As_t$ es el porcentaje de la población asalariada a tiempo t.
f es el factor de reincorporación a la vida laboral.
x es el grupo de edad ($x \in \{2, ..., 7\}$)
$\bar{x}$ es la edad promedio del grupo de edad x.
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

Al calcular de esta manera la pensión por invalidez no se refleja de manera adecuada las pensiones otorgadas por el IMSS. Para corregir lo anterior se multiplicó por el porcentaje de pensiones otorgadas del Instituto Mexicano del Seguro Social ($\%PO_t$). Los datos se obtuvieron de las memorias estadísticas del IMSS.

Serie	Modelo	Parámetros
$\%PO_t$	Regresión	$\hat{\beta}_0 = -1.2570$; $\hat{\beta}_1 = -0.0006$

Nótese que dicho factor no está dividido por grupo de edad o sexo, en consecuencia el producto del factor se hará sobre el total de la pensión de invalidez.

Nótese además que al multiplicar *el número esperado de personas asalariadas con cáncer al año t* ($[{}_xCEI_t^j][\,\%As_t^j]$) por ($1 - {}_xf^j$) se esta calculando *el número esperado de asalariados con cáncer que no regresarán a la vida laboral.*

El factor de reincoorporación a la vida laboral (f) fue obtenido de un estudio de origen cubano (*Bibliografía:* **[6]** y **[7]**).

Una vez más, sólo consideramos a las población asalariada.

3.3.4. Costo de oportunidad del cuidador.

El cáncer es una padecimiento que el enfermo no puede sobrellevar solo, en partícular durante los primeros días de tratamiento, ya que los mismos son muy agresivos con el cuerpo. Por esto, el enfermo casi siempre va acompañado de algún familiar cercano, mismo que deja sus actividades cotidianas, entre ellas las laborales por acompañar al enfermo. En consecuencia, deja de percibir un ingreso monetario. El costo de oportunidad del cuidador se define como el ingreso que dejará de recibir el cuidador del enfermo a causa de los cuidados de este.

Por tanto, el costo de oportunidad del cuidador se calcula como:

$$ {}_{x}COC_{t}^{j} = 270 \; {}_{x}CEI_{t}^{j} \; {}_{y^*}ID_{t}^{j} $$

donde:

COC_t es el costo de oportunidad del cuidador a tiempo t.
270 es el número promedio de incapacidad por cáncer.
CEI_t son los casos esperados incidentes de cáncer al tiempo t.
ID_t es el ingreso diario del cuidador a tiempo t.
x es el grupo de edad del enfermo ($x \in \{1, ..., 10\}$)
y^* es el grupo de edad del cuidador ($y^* \in \{1^*, ..., 10^*\}$)
j es el sexo ($j = 1 \rightarrow$ hombre, $j = 2 \rightarrow$ mujer).

En este caso x y y^*, no necesariamente son el mismo grupo de edad. Para obtener el costo de oportunidad del cuidador, se supuso que todos los enfermos de cáncer tienen un solo cuidador directo y se simularon dos escenarios:

1. Usando el SBC del IMSS. En este caso solo se simuló si el cuidador sería hombre o mujer (suponiendo que son equiprobables) ya que el SBC no depende del grupo de edad.

2. Usando el ingreso anual reportado por la ENOE. Para este caso se simuló tanto el sexo como el grupo de edad, ya que los datos reportados por la ENOE sí dependen de la edad (asumiendo que los diez grupos de edad y los dos sexos son equiprobables).

Capítulo 4

Resultados.

Se estima que para el año 2014, en México habrán 605,758 casos de cáncer ($IC_{95\,\%} = (540,665.00, 670,852.00)$) de los cuales 167,534 serán casos nuevos. Además el 41.46 % es de los casos prevalentes corresponde a hombres ($IC_{95\,\%} = (41.91,\ 48.01)$). Si se continúa como hasta hoy en día, el número de casos prevalentes en cáncer podrían duplicarse para el año 2020, con un total de 1,262,861 ($IC_{95\,\%} = (1,079,419.00, 1,446,303.00)$), de los cuales 216,679 serían casos nuevos. Una vez más, el porcentaje mayor lo presentan las mujeres (57.33 % con un $IC_{95\,\%} = (51.62,\ 62.92)$).

Figura 4.1: Casos incidentes esperados en cáncer.

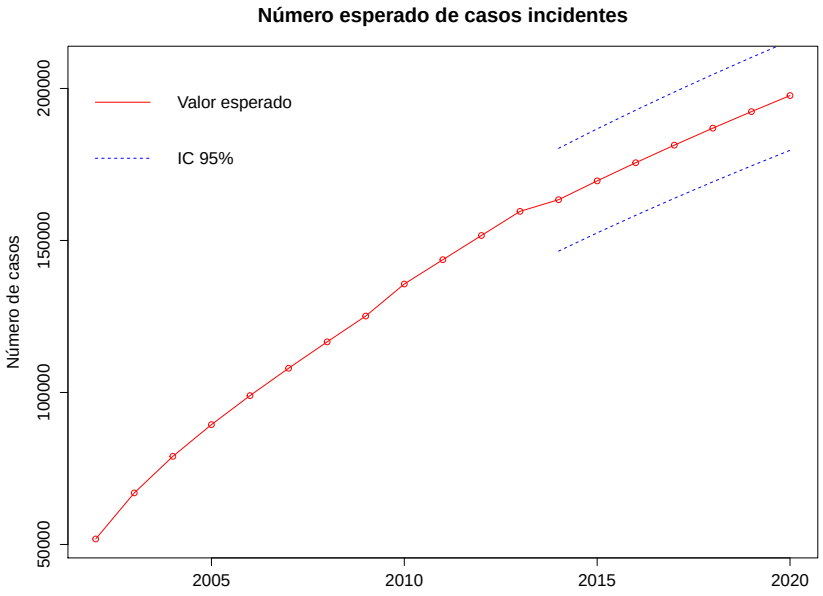
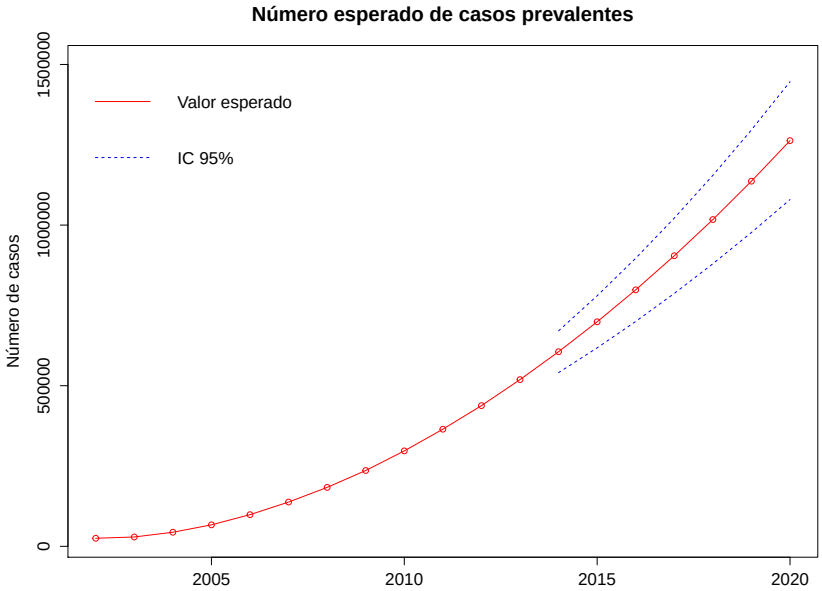


Figura 4.2: Casos prevalentes esperados en cáncer.



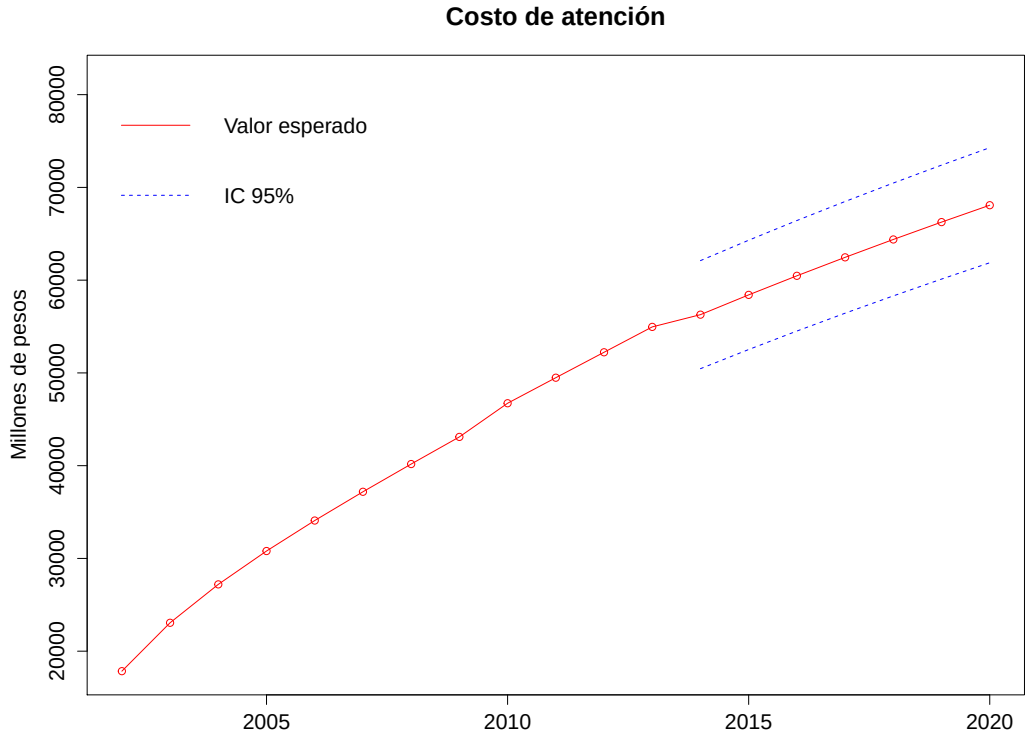
Se debe de tener cuidado con la interpretación de dicho resultado, ya que aunque pudiése advertirse un crecimiento geométrico en el número de casos prevalentes y esto pudiera generar polémica al respecto, se debe recordar que los casos esperados prevalentes fueron definidos como:

$$ {}_xCEP_t^j = {}_xCEI_t^j - {}_xDc_t^j + {}_xCEP_{t-1}^j $$

Donde notemos que existe una dependencia en el tiempo con el dato anterior. Entonces no es erróneo suponer un crecimiento aritmético o geométrico para los casos prevalentes, teniendo en cuenta que dicho efecto es causado por *los casos incidentes (a tiempo t) que no murieron durante ese año y se convirtieron en casos prevalentes para el año siguiente* $(t+1)$.

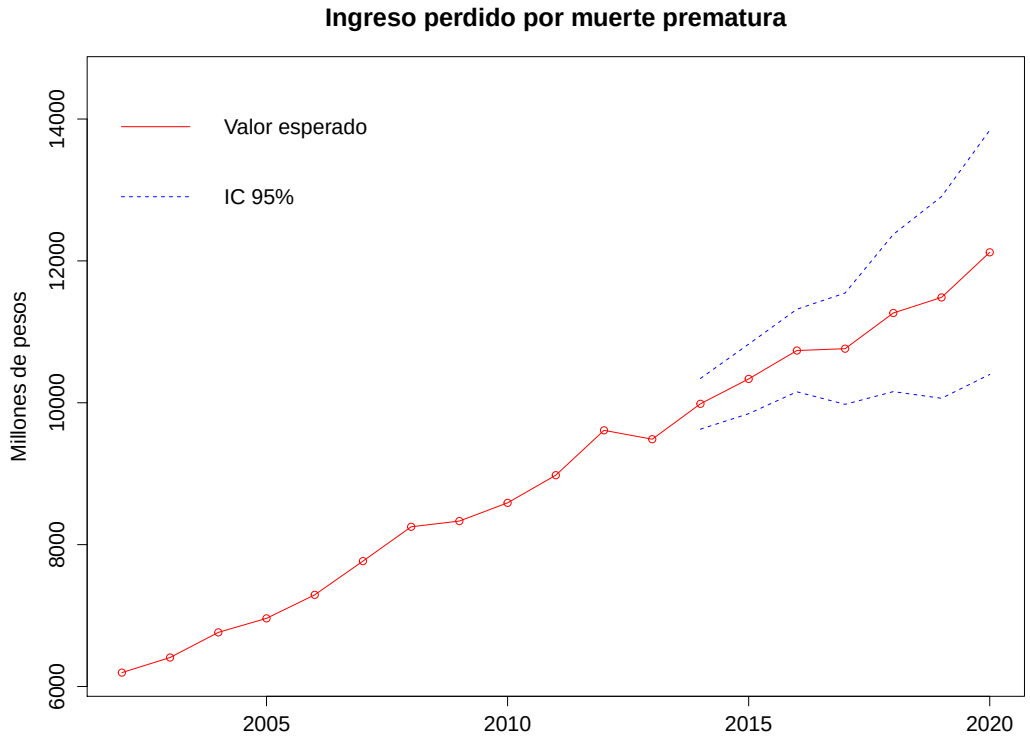
Por otro lado, el estimado del costo de atención anual es de $\$344,421.26$ ($IC_{95\,\%} = (237,182.76,\ 451,659.77)$) y en consecuencia, la estimación del costo de atención para los casos incidentes al año 2014 será de 56,280.82 millones de pesos ($IC_{95\,\%} = (55,455.41,\ 57,106.22)$) y para el año 2020 será de 68,079.62 millones de pesos ($IC_{95\,\%} = (66,877.25,\ 69,281.98)$).

Figura 4.3: Costo de atención por cáncer en México.



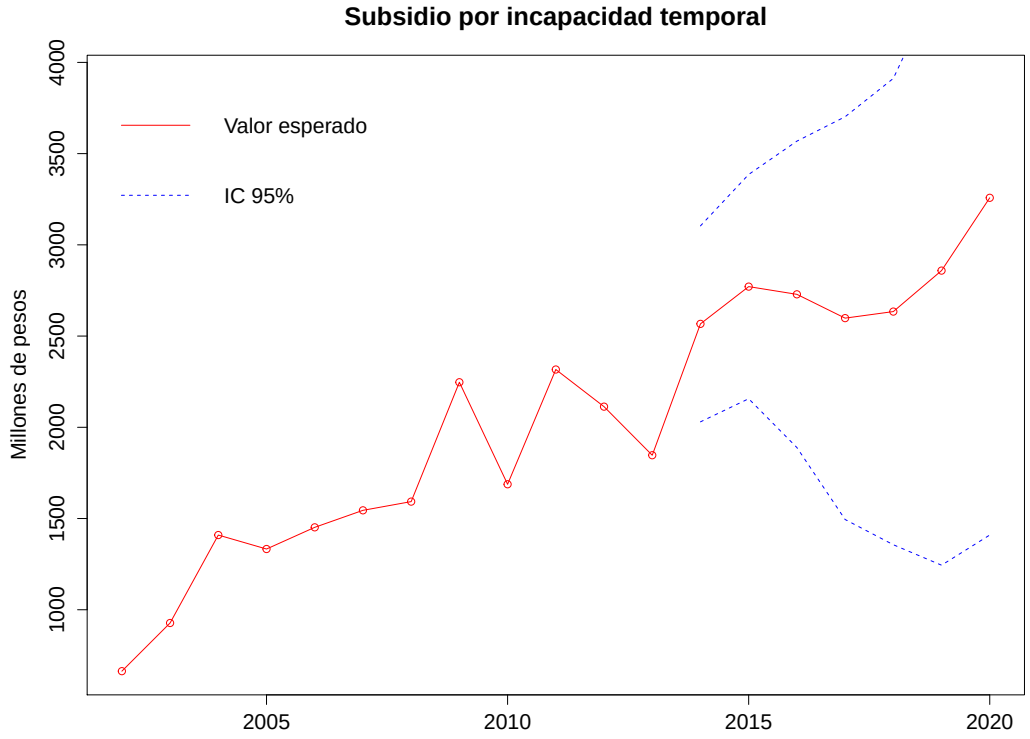
La estimación del ingreso perdido por muerte prematura es de $12,041.88$ millones de pesos ($IC_{95\,\%} = (11,218.55,\ 12,453.54)$ utilizando el SBC del SEM. Y de $7,928.73$ millones de pesos ($IC_{95\,\%} = (7,321.39,\ 8,232.41)$ utilizando los datos de la ENOE, ambos al año 2014. Usaremos el promedio de ambos como el estimador del ingreso perdido por muerte prematura, cuyo valor es de $9,985.31$ millones de pesos, cuya respectiva proyección al año 2020 es de $12,120.91$ millones de pesos. El mayor porcentaje de participación lo representan los hombres con un $66.78\,\%$ del total ($IC_{95\,\%} = (63.57,\ 70.00)$).

Figura 4.4: Número total de casos prevalentes esperados en cáncer.



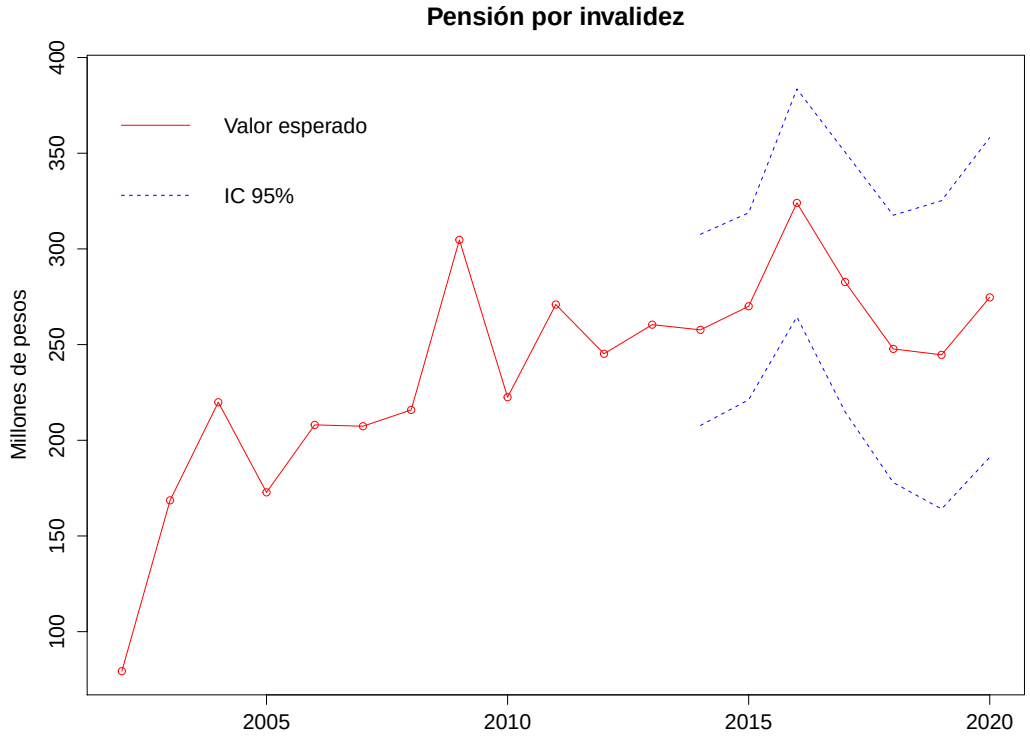
La estimación del subsidio por incapacidad temporal es de $2,556.87$ millones de pesos ($IC_{95\,\%} = (3,103.91,\ 2,029.82)$) al año 2014 y cuya respectiva proyección al año 2020 es de $3,257.88$ millones de pesos ($IC_{95\,\%} = (1,409.53,\ 5,106.21)$). En este caso, la participación es dominada por las mujeres con un porcentaje estimado del $54.56\,\%$ del total ($IC_{95\,\%} = (51.22,\ 57.90)$).

Figura 4.5: Subsidio por incapacidad temporal.



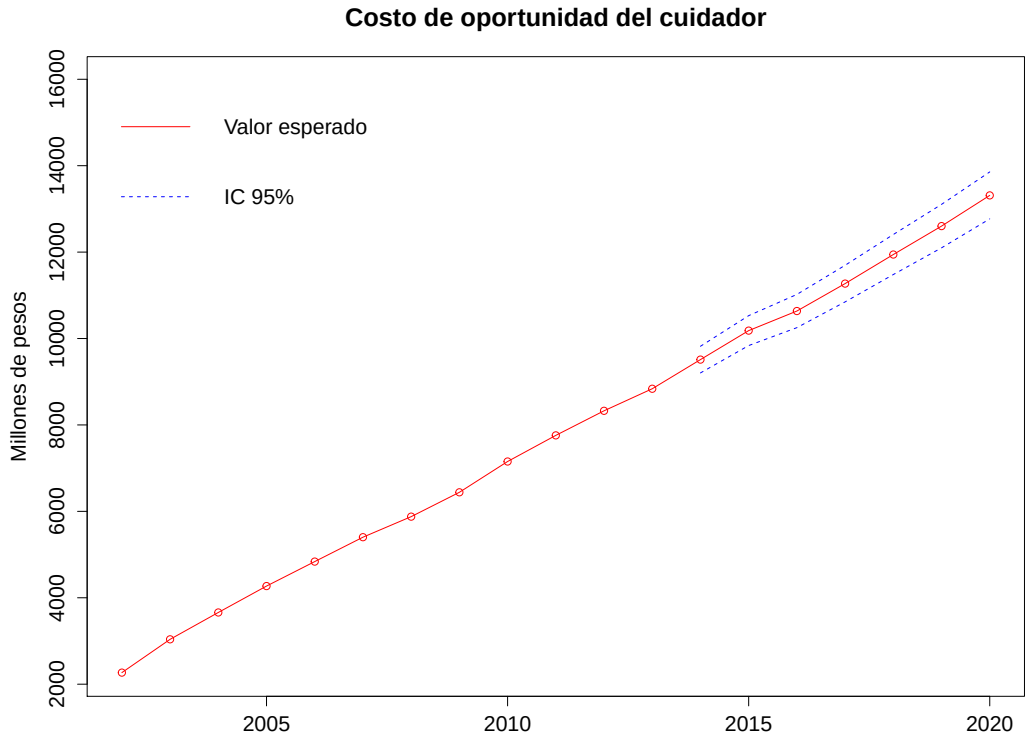
La estimación del costo por concepto de pensión por invalidez en el 2014, es de 257.71 millones de pesos ($IC_{95\,\%} = (207.74,\ 307.68)$) cuya respectiva proyección al año 2020 es de 274.65 millones de pesos ($IC_{95\,\%} = (191.14,\ 358.16)$). Una vez más, la participación es dominada por las mujeres con un porcentaje estimado del 62.30 % del total ($IC_{95\,\%} = (59.00,\ 65.60)$).

Figura 4.6: Pensión por invalidez.



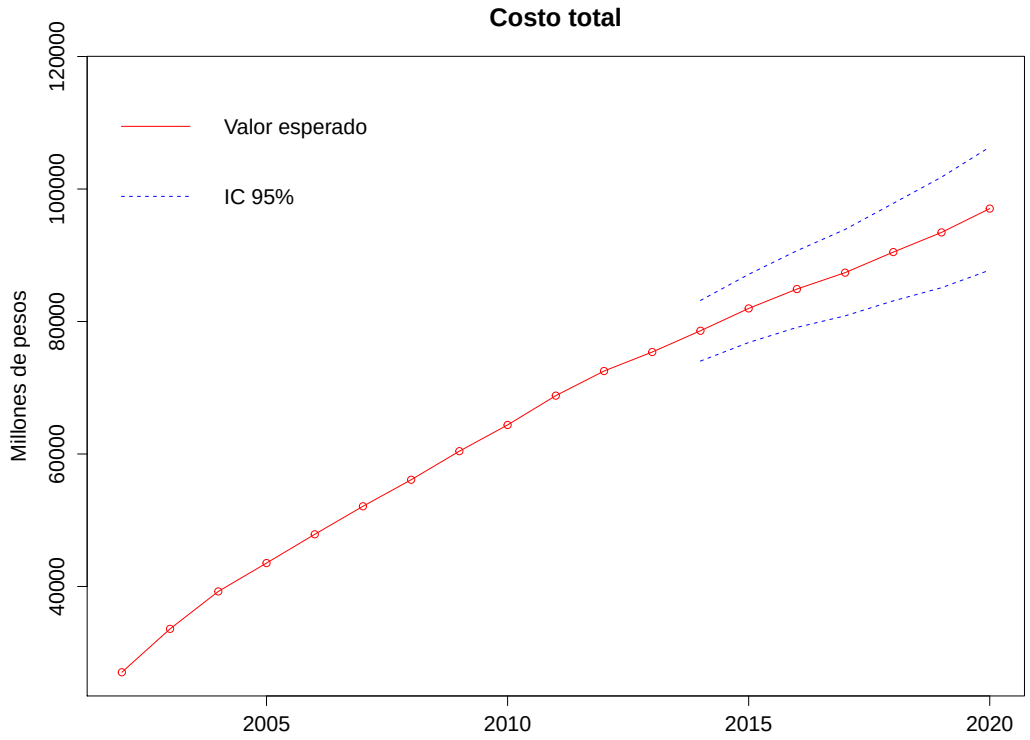
Por último, la estimación del costo de oportunidad del cuidador en el 2014, es de $12,166.38$ millones de pesos ($IC_{95\,\%} = (11,811.42,\ 12,521.54)$ utilizando el SBC del SEM. Y de $6,858.20$ millones de pesos ($IC_{95\,\%} = (6,596.34,\ 7,120.05)$) utilizando los datos de la ENOE. Una vez más, usaremos el promedio de ambos como el estimador del costo de oportunidad del cuidador, cuyo valor es de $9,512.29$, cuya respectiva proyección al año 2020 es de $13,312.39$. El mayor porcentaje de participación lo representan las mujeres con un $54.21\,\%$ del total ($IC_{95\,\%} = (50.22,$ $57.17)$).

Figura 4.7: Costo de oportunidad del cuidador.



Por lo tanto, se estima un total de 1,262,861 de casos prevalentes en cáncer al año 2020, los cuales implicarán un costo total de 97,045.44 millones de pesos, de los cuales el 71.08 % corresponden al costo de atención de los pacientes.

Figura 4.8: Costo total por cáncer en México.



Capítulo 5

Conclusiones.

Se estima que para el año 2014, en México habrán 605,758 casos de cáncer y se espera que para el año 2020 se tengan un total de 1,262,861. Así mismo, se espera un porcentaje menor de casos de cáncer en hombres que en mujeres esto puede atribuirse principalmente a la tasa mortalidad ya que es mayor para hombres que para mujeres, en consecuencia, aunque hayan mayor número de casos incidentes en hombres que mujeres, estos fallecen más rápido, lo cual provoca que haya más casos prevalentes de mujeres con cáncer.

Por otro lado, y respondiendo a la segunda pregunta objetivo[1], el monto necesario para atender a todos los enfermos con cáncer al año 2014 es de 78,602.98 millones de pesos. Al compararlo con el presupuesto federal destinado al sector salud al año 2014 (350,361.64 millones de pesos [2]) nos damos cuenta de que el costo de atención por cáncer representa cerca del 22.43 % sin mencionar que existen otras enfermedades importantes que aquejan a la población mexicana como las enfermedades cardiovasculares, diabetes y enfermedades respiratorias.

Cabe mencionar que no todos los rubros del costo indirecto afectan al sector público. Específicamente solo el subsidio por incapacidad temporal y la pensión por invalidez van a cargo del IMSS, entonces el costo indirecto generado por cáncer en México al año 2014 fue de $2,824.87$ millones de pesos. Además, del total de casos incidentes, 115,483 corresponden a personas asalariadas, lo cual implica un costo de atención de 39,898.79 millones de pesos. Por tanto, el costo total por cáncer para el IMSS al año 2014 es de 42,723.67 millones de pesos. Al compararlo con el presupuesto federal destinado para el IMSS al año 2014 (217,312.79 millones de pesos[1]), nos damos cuenta de que el costo de atención representa el 19.66 % del presupuesto. De aquí podemos notar, una vez más, que el presupuesto no es suficiente si se consideran otras enfermedades.

Ante estos resultados y dadas las proyecciones y pronósticos realizados, tanto

[1] Véase Introducción.

[2] Según el Presupuesto de Egresos de la Federación para el Ejercicio Fiscal 2014 *Diario Oficial de la Federación 03-12-2013*

del costo total como del número de casos prevalentes en cáncer, podemos considerar dos posibles soluciones:

- **Asignar más recursos al sector salud**. Tendría que aumentar al menos en un 100.00 % para que el costo por cáncer represente cerca del 10 % del presupuesto asignado al IMSS (el cual aún es un porcentaje demasiado alto) y poder atender a todos los casos esperados de cáncer así como otras enfermedades importantes.

- **Implementar políticas para el control de los principales factores de riesgo** como lo son: el tabaquimo, el alcoholismo, la obesidad y la mala nutrición ya que juntos asocian cerca del 30 % de los casos de cáncer.

La alternativa ideal sería implementar ambas soluciones para atender los casos actuales y prevenir casos futuros. Sin embargo, dichas soluciones son a largo plazo, ya que para poder tener resultados notorios al implementar políticas de control deben pasar aproximadamente 10 años.

Cabe mencionar, una vez más, que el trabajo está basado en datos nacionales y estimaciones internacionales, en consecuencia se debe tener en consideración que la veracidad de los resultados presentados depende de la certeza de los datos ocupados durante en mismo, por lo tanto no se rechaza la posibilidad de tener una variación en los resultados. Así mismo, se debe tener en cuenta que el tiempo de las proyecciones y pronósticos realizados es un tiempo demasiado largo, pronosticar a esta temporalidad es demasiado ambicioso ya que si en un año pueden pasar demasiados acontecimientos que modifiquen el estudio ¿qué se puede esperar dentro de 8 años?. De igual forma, los modelos estadísticos empleados para estimaciones realizadas dependen tanto de la calidad de la información, como de la cantidad de datos. Sin embargo, ahora se tiene una primer aproximación a nivel nacional de la magnitud que representa el cáncer para la salud pública y se pueden tomar medidas de prevención y control de los factores de riesgo en base a estos resultados, el cual fue el objetivo de este trabajo y la mayor aportación del mismo, ya que ahora se cuenta con estimaciones sobre el número de casos de cáncer en México y su impacto económico.

Por último, la principal forma de reducir la probabilidad de desarrollar cáncer es tener conciencia de los factores de riesgo a los cuales nos exponemos diariamente (tales como tabaco, alcohol, un mala nutrición, una vida sedentaria, etcétera) así como la severidad y frecuencia, para evitarlos en medida de lo posible o bien exponerse de manera responsable y así no tener que sufrir las consecuencias de una vida de excesos en un futuro no muy lejano.

Bibliografía.

[**1**] Brockwell P.J. and Davis R.A. *Introduction to Time Series and Forecasting.* New York: Springer-Verlag, 1996.

[**2**] Montgomery D.C., Peck E.A., Vining G.G. *Introduction to Linear Regression Analysis.* Wiley & Sons, 2012.

[**3**] Guerrero, V.M. (1993). *Análisis Estadístico de Series de Tiempo Económicas.* México: UAM-Iztapalapa.

[**4**] IARC. (2014). *Cáncer*. enero 8, 2015, de IARC Sitio web:
http://www.cancer.gov/espanol/cancer/que-es

[**5**] GLOBOCAN. (2012). *Tasas de Incidencia en Cáncer México*. noviembre 4, 2014, de IARC Sitio web:
$http://globocan.iarc.fr/Pages/fact\_sheets\_population.aspx$

[**6**] Domínguez Alonso, Emma; H. Seuc. *Esperanza de vida ajustada por algunas enfermedades crónicas no transmisibles*. Rev Cubana de Hig y Epidemiol 2005
http://bvs.sld.cu/revistas/hie/vol41200/hie05301.htm

[**7**] Seuc AH, Domínguez E, Galán Y. *Esperanza de vida ajustada por Cáncer*. Rev Cubana Hig y Epidemiol 2003;41(2).
http://bvs.sld.cu/revistas/hie/vol41200/hie05203.htm

[**8**] The World Health Report. World Health Organization: Geneva; 2013. http://www.who.int/topics/worldhealtreport/es/

[**9**] Chen HS, Portier K, Ghosh K, et al. *Predicting US- and state-level cancer counts for the current calendar year: Part I: evaluation of temporal projection methods for mortality.* Cancer. 2012; 118: 1091-1099.

[**10**] Zhu L, Pickle LW, Ghosh K, et al. *Predicting US- and state-level cancer*

counts for the current calendar year: Part II: evaluation of spatiotemporal projection methods for incidence. Cancer. 2012; 118: 1100-1109.

[**11**] Angela B. Mariotto, K. Robin Yabroff, Yongwu Shao, Eric J. Feuer, Martin L. Brown. *Projections of the Cost of Cancer Care in the United States: 2010–2020.* Journal of the National Cancer Institute. 2011.

[**12**] Weiss W. *Cigarette smoking and lung cancer trends. A light at the end of the tunnel?* Chest. 1997; 111: 1414-1416.

[**13**] Jemal A, Thun MJ, Ries LA, et al. *Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control.* J Natl Cancer Inst. 2008; 100: 1672-1694.

[**14**] WHO. (2014). *Epidemiología.* febrero 3, 2015, de WHO Sitio web: http://www.who.int/topics/ep

[**15**] Ferlay, J., Shin, H.-R, Bray, F., Forman, D., Mathers, C. and Parkin, D. M. (2010) *Estimates of the worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int. J. Cancer, 127:2893-2817

[**16**] *Manual CTO de Medicina y Cirugía*, Tomo I, ISBN 84-930264-3-3.

[**17**] $http://www.infocancer.org.mx/mortalidad-tumores-malignos-en-la-poblacin-de-20-aos-y-ms-con788i0.html\#sthash.IW3qEIYg.dpuf$

Apéndice A

Definicinoes extras.

La epidemiología requiere de herramientas no solo matemáticas, sino también económicas y demográficas para su estudio e interpretación. En este capítulo se darán las definiciones que ayudaran a comprender de manera clara el presente trabajo.

A.1. Demográficas.

***Definición* A.1** *Los censos son métodos estadísticos que se emplean para poder conocer las características de los habitantes de México y sus viviendas a nivel nacional, estatal, municipal, por localidad, por grupos de manzanas y hasta por manzana.*

En México el Censo se realiza cada diez años, en aquéllos terminados en cero; y el Conteo, cada diez años también, pero en aquéllos terminados en cinco.

***Definición* A.2** *La tasa de mortalidad general es la proporción de personas que fallecen respecto al total de la población.*
Se calcula cómo:

$$m = 100,000 * \frac{D}{T}$$

donde:
m *es la tasa de mortalidad.*
D *son los fallecimientos.*
T *es el total de la población.*

***Definición* A.3** *La tasa de mortalidad específica se refiere a la proporción de personas con una característica particular que mueren respecto al total de personas que tienen esa característica, puede ser por edad, sexo o alguna otra característica*

propia de la población de estudio.
Se calcula cómo

$$m_j = 100,000 * \frac{D_j}{T_j}$$

donde:
m_j *es la tasa de mortalidad por la causa específica* j.
D_j *son los fallecimientos por la causa* j.
T_j *es el total de la población que puede fallecer por la causa* j.

Con base a la definición anterior es facíl deducir la siguiente definición.

***Definición* A.4** *La tasa de incidencia es la proporción de personas que desarrollan una enfermedad* x *respecto al total de personas que podría desarrollar dicha enfermedad.*
Se calcula cómo

$$m_x = 100,000 * \frac{M_x}{T}$$

donde:
m_x *es la tasa de incidencia de la enfermedad* x.
M_x *es el número de personas que desarrollaron la enfermedad* x.
T *es el total de la población.*

Nótese que a diferencia de la definicion 3.2, se está dividiendo entre el total de población, ya que salvo algunos casos muy especiales, se supone que cualquier miembro de la población puede desarrollar cualquier enfermedad.

Las tasas anteriormente calculadas pueden ser delimitadas por algún periodo de tiempo t, donde t puede ser medido en horas, días, meses, años, etcétera. De ahora en adelante, cada vez que se hable de *tasas* (sin importar si son de mortalidad o incidencia) se supondrá que la temporalidad es anual.

A.2. Económicas.

***Definición* A.5** *La Población Económicamente Activa (PEA) es el conjunto de personas de más de 15 años que desempeñan una ocupación, o bien, si no la tienen, la buscan activamente. Ello excluye a los pensionados y jubilados, a las amas de casa, estudiantes y rentistas así como, por supuesto, a los menores de edad.*

***Definición* A.6** *La Población Económicamente Activa y Ocupada (PEAO) es el conjunto de personas de más de 15 años que desempeñan una ocupación.*

***Definición* A.7** *Deflactar es transformar una magnitud económica expresada en términos monetarios a precios corrientes en otra magnitud expresada también en términos monetarios, pero a precios del año cero o año base, al objeto de eliminar del valor de dicha magnitud el efecto de la inflación o subida de precios mediante un índice de precios que actúa como deflactor.*

En México, se utiliza el $INPC$ para deflactar mediante la siguiente fórmula.

$$K_t = K_h \left(\frac{INPC_h}{INPC_t} \right)$$

donde:

K_t es la cantidad deflactada a tiempo t.
K_h es la cantidad a deflactar que esta a tiempo h.
$INPC_i$ es el índice nacional de precios al consumidor a tiempo i.

***Definición* A.8** *Según la Ley del Seguro Social, en su artículo 27, el salario base de cotizacion se integra con los pagos hechos en efectivo por cuota diaria y las gratificaciones, percepciones, alimentacion, habitacion, primas, comisiones, prestaciones en especie y cualquier otra cantidad o prestacion que se entregue al trabajador por sus servicios.*

***Definición* A.9** *Un subsidio es una prestación pública asistencial de carácter económico y de duración determinada.*

A.3. Actuariales.

***Definición* A.10** *Dado una edad x, definimos a*

- *l_x como el número de personas vivas a edad x.*
- *d_x como el número de personas que murieron entre edad x y $x+1$.*

A l_0 se le conoce como *rádix* y generalmente se presenta en múltiplos de 100.

***Definición* A.11** *Definimos a la probabilidad de supervivencia a t años de una persona de edad x como:*

$$ {}_tp_x = \frac{l_{x+t}}{l_x} $$

En consecuancia de la definición anterior, la probabilidad de muerte entre edades x y $x+t$ se calcula cómo:

$$ {}_tq_x = 1 - {}_tp_x $$

***Definición* A.12** *Dada una tasa de interés i y un tiempo n, el valor presente se define como:*

$$V^n = (1+i)^{-n}$$

El *valor presente* se interpreta como *el valor al día de hoy de una unidad monetaria pagadera a tiempo n.*

***Definición* A.13** *Una anualidad anticipada temporal a n años para una persona de edad x y una tasa de interés i se define como:*

$$\ddot{a}_{x:\overline{n}|} = \sum_{t=0}^{n} V^t {}_tp_x$$

***Definición* A.14** *Una anualidad anticipada vitalicia para una persona de edad x y una tasa de interés i se define como:*

$$\ddot{a}_x = \sum_{t=0}^{\infty} V^t {}_tp_x$$

Apéndice B

Gráficas.

Aquí se presentan las graficas de los modelos utilizados. Las gráficas se presentarán en el orden en como aparecieron en el **capítulo 4.**

Figura B.1: Tasas de incidencia en cáncer por grupos de edad para hombres

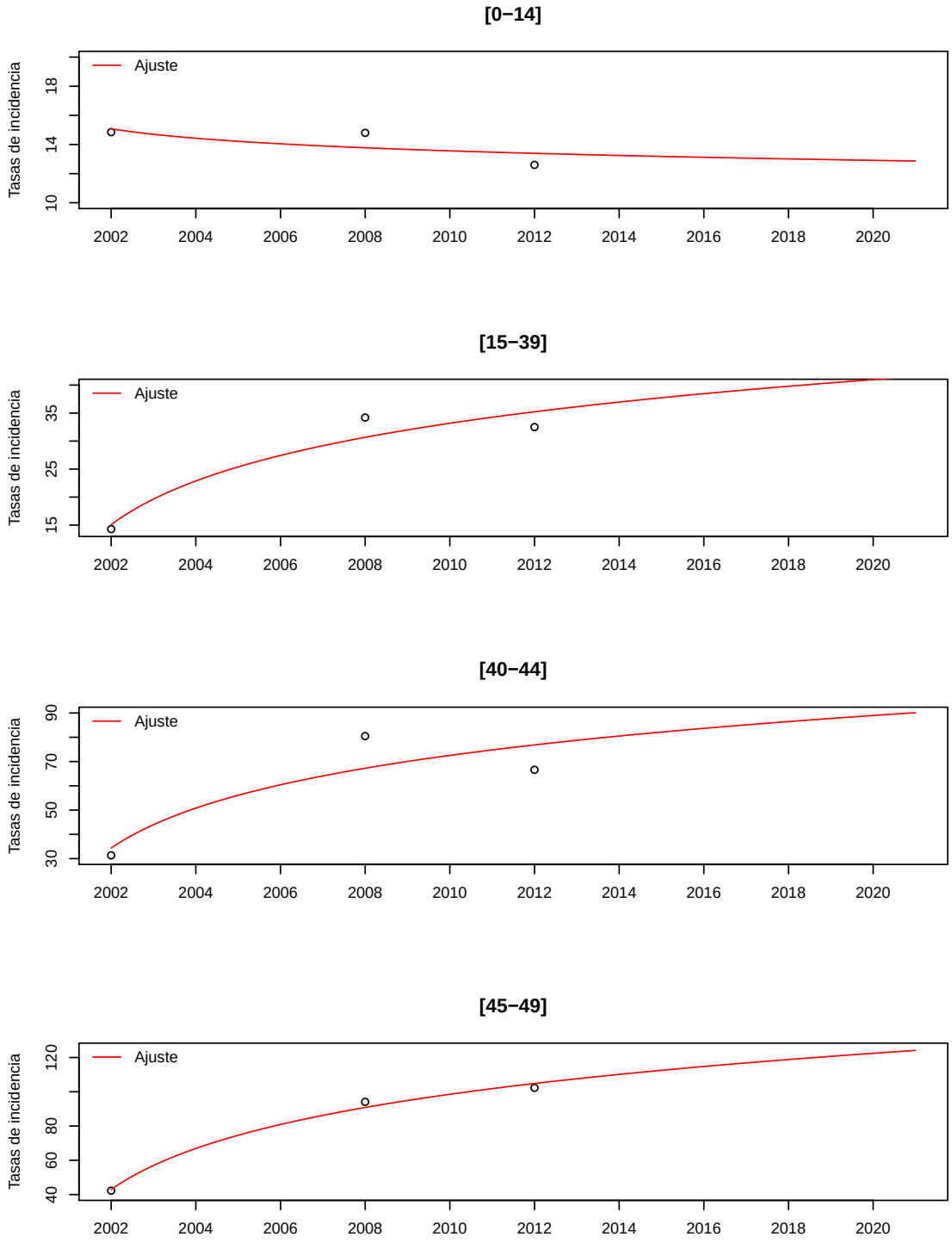


Figura B.2: Tasas de incidencia en cáncer por grupos de edad para hombres

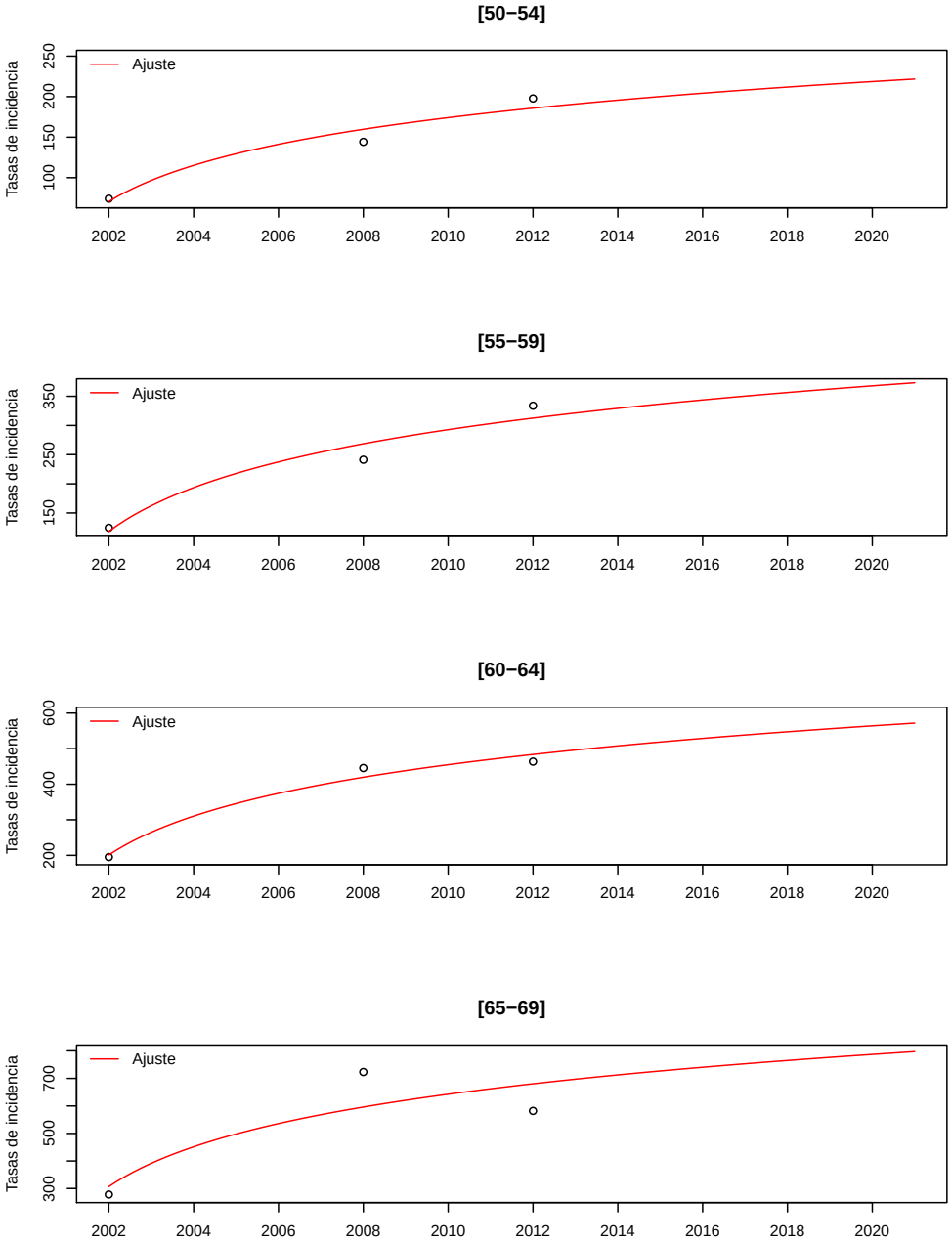


Figura B.3: Tasas de incidencia en cáncer por grupos de edad para hombres

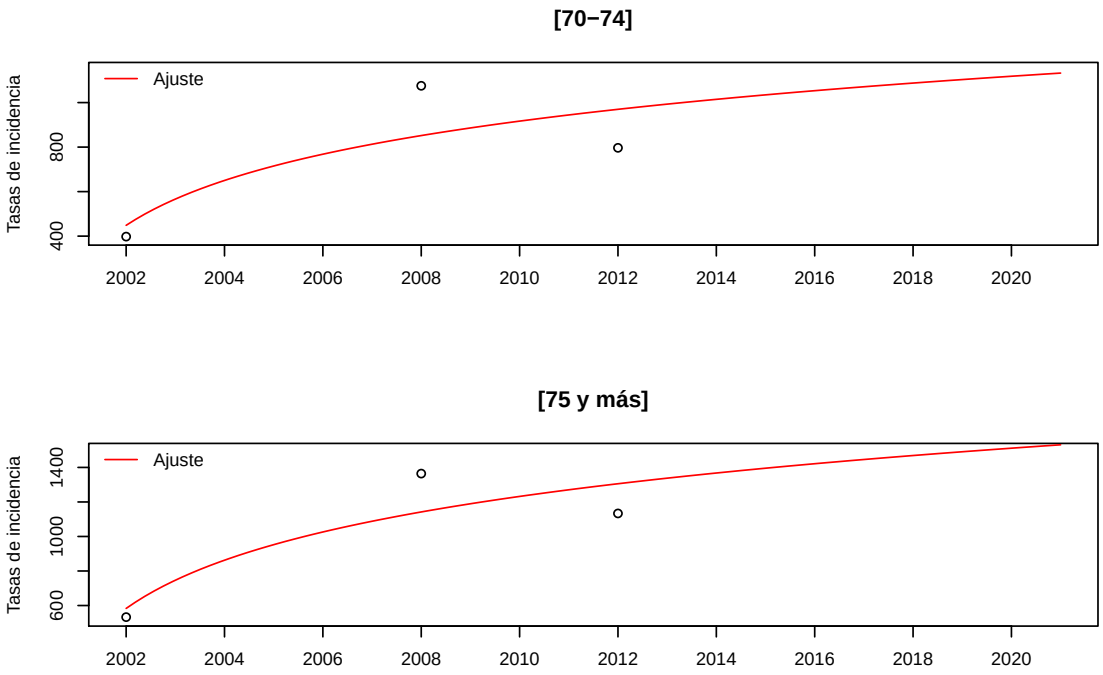


Figura B.4: Tasas de incidencia en cáncer por grupos de edad para mujeres

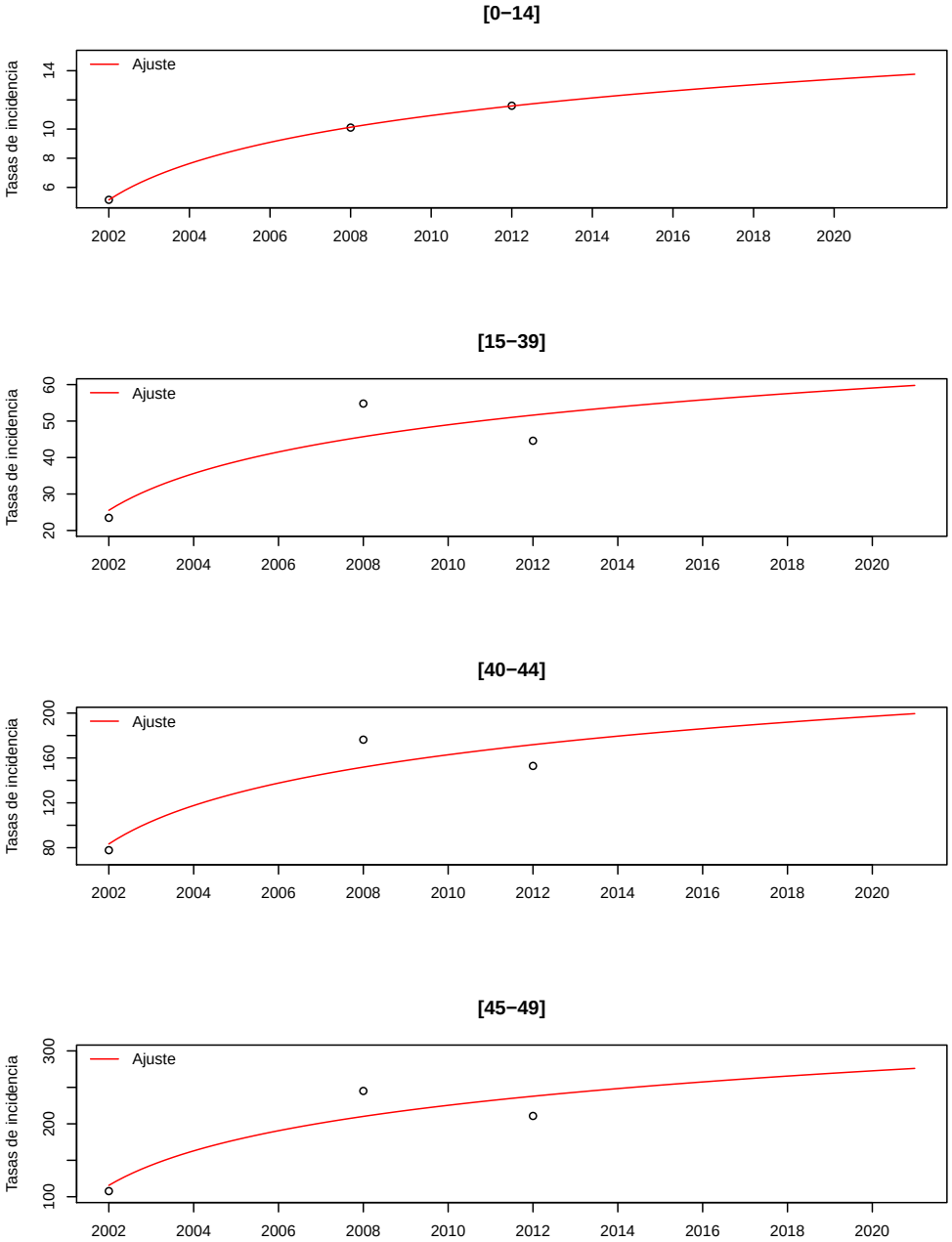


Figura B.5: Tasas de incidencia en cáncer por grupos de edad para mujeres

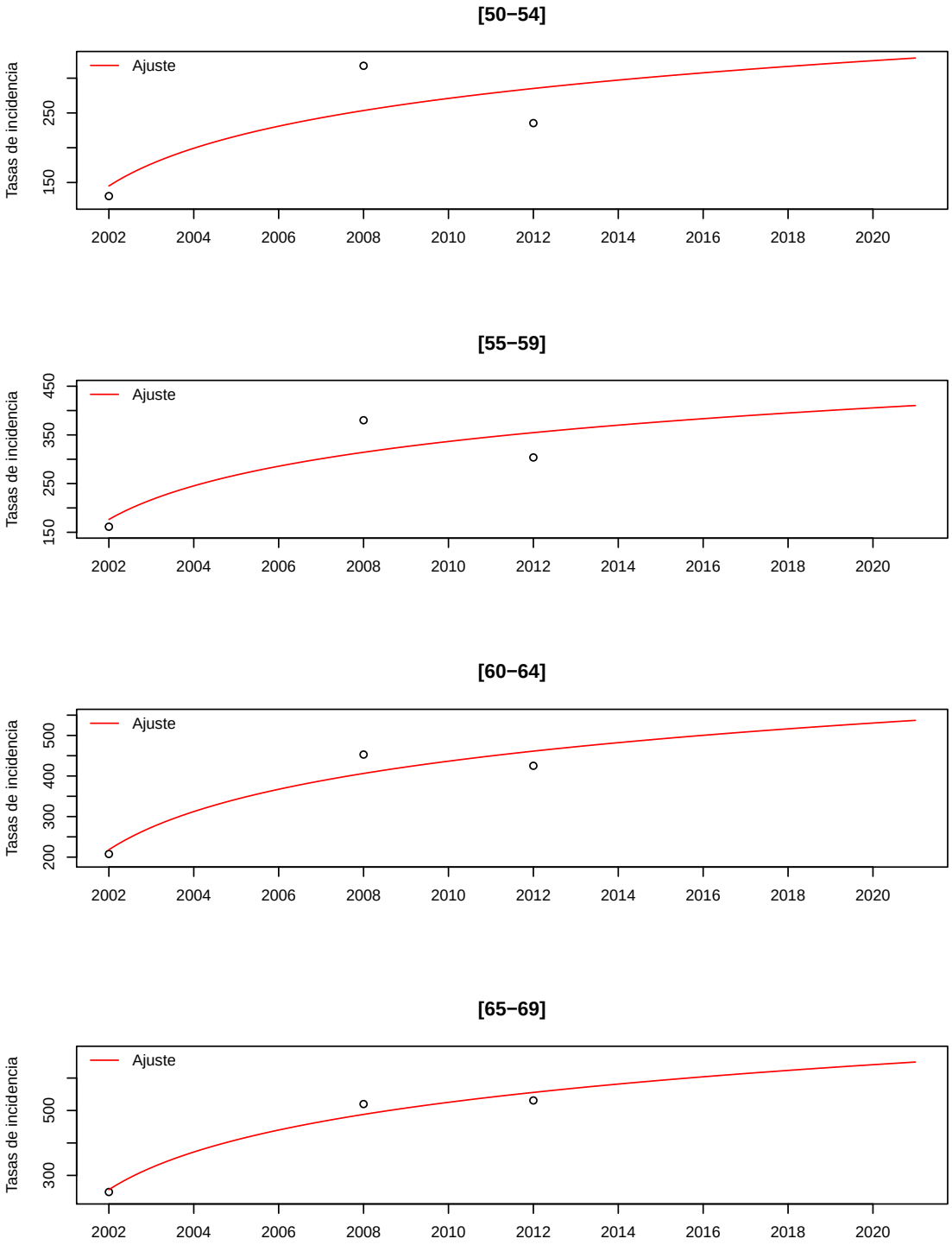


Figura B.6: Tasas de incidencia en cáncer por grupos de edad para mujeres

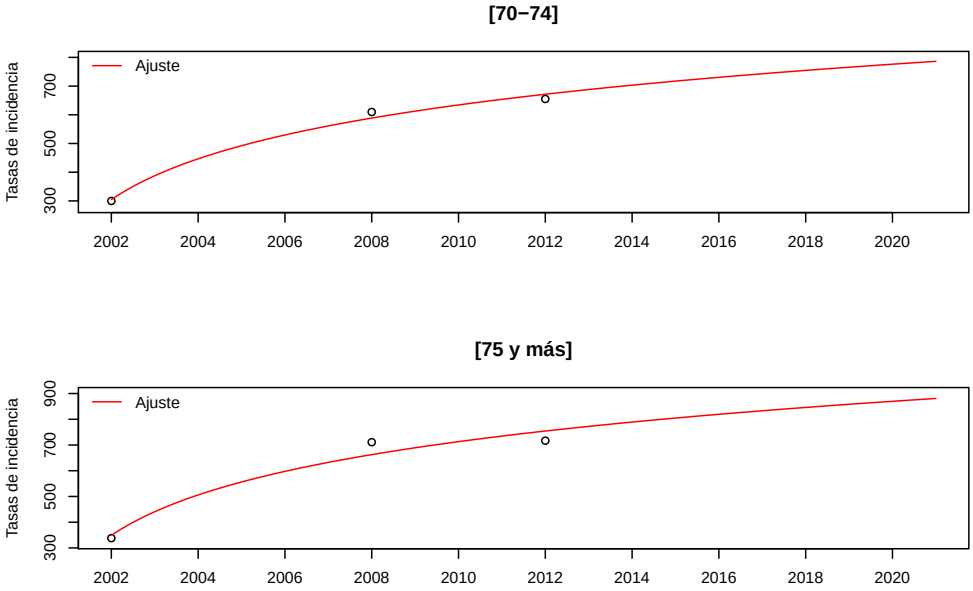


Figura B.7: Defunciones por cáncer por grupos de edad para hombres

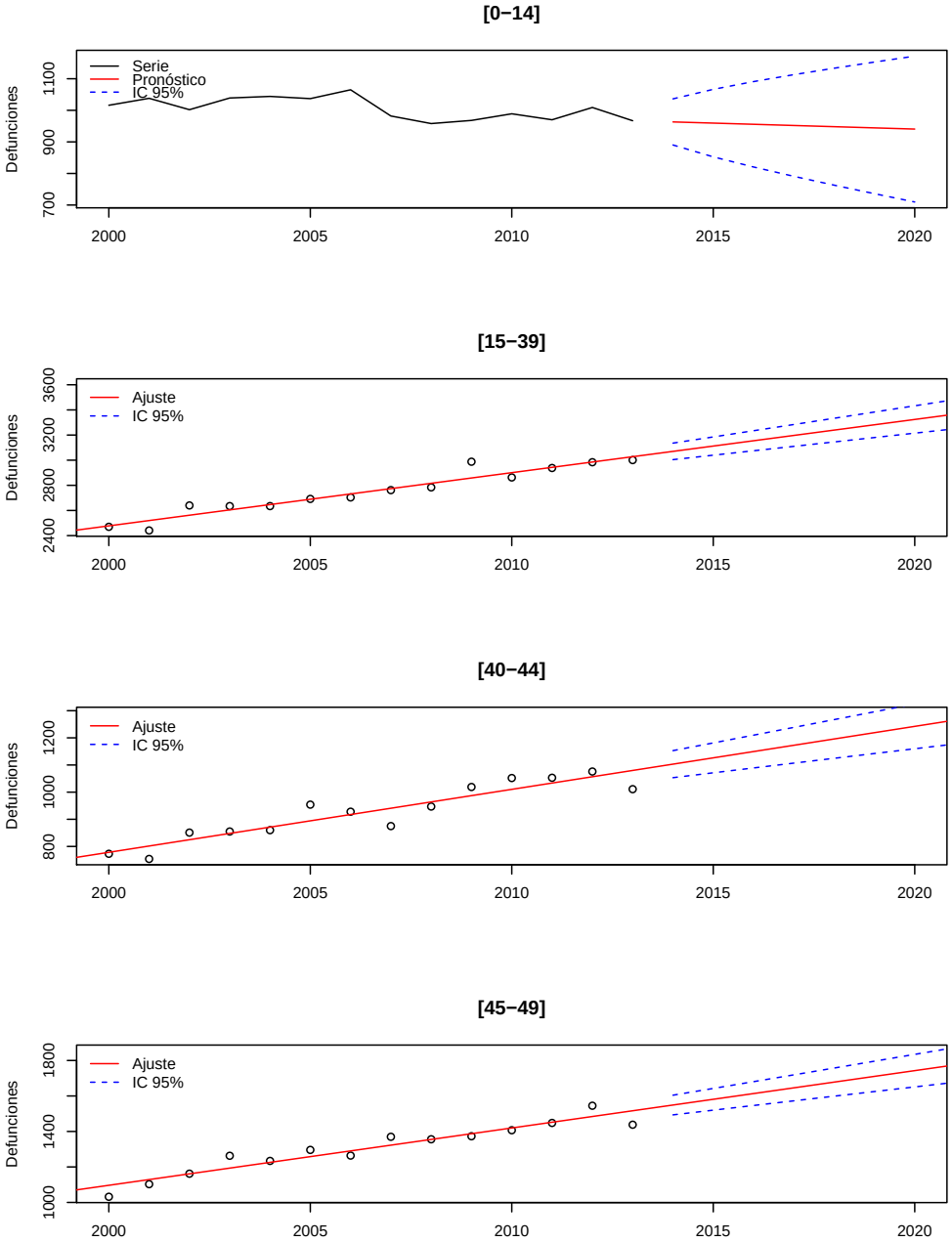


Figura B.8: Defunciones por cáncer por grupos de edad para hombres

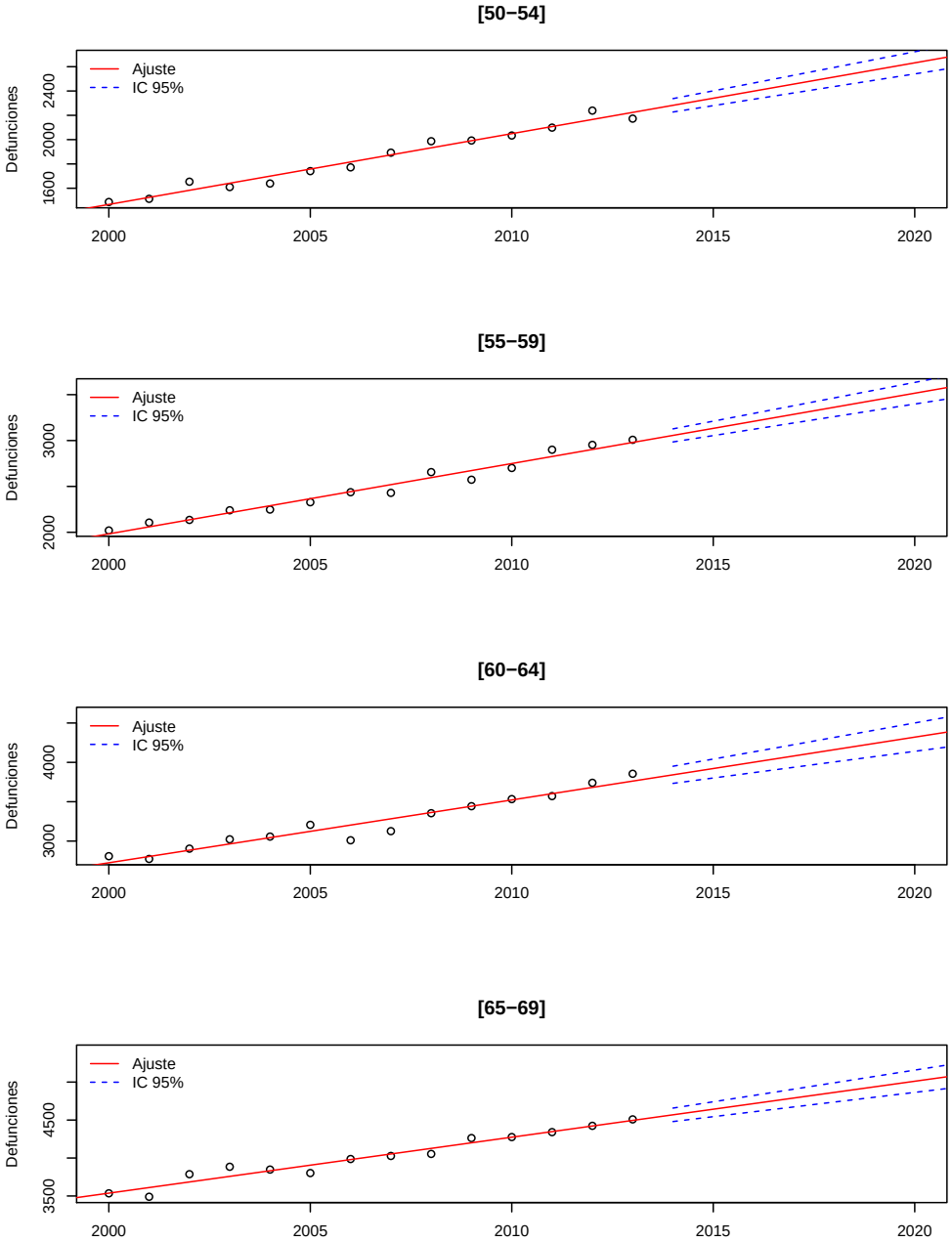


Figura B.9: Defunciones por cáncer por grupos de edad para hombres

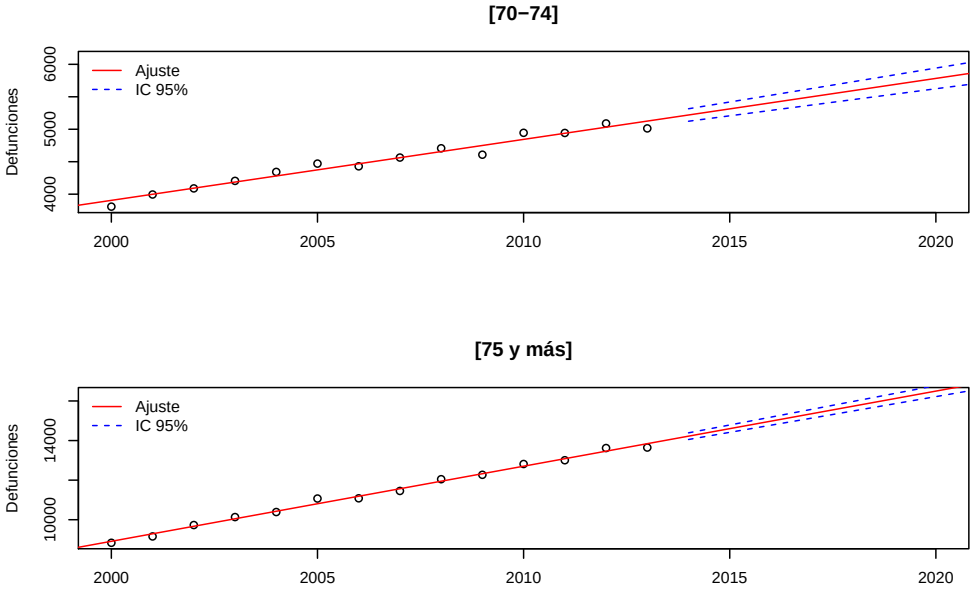


Figura B.10: Defunciones por cáncer por grupos de edad para mujeres

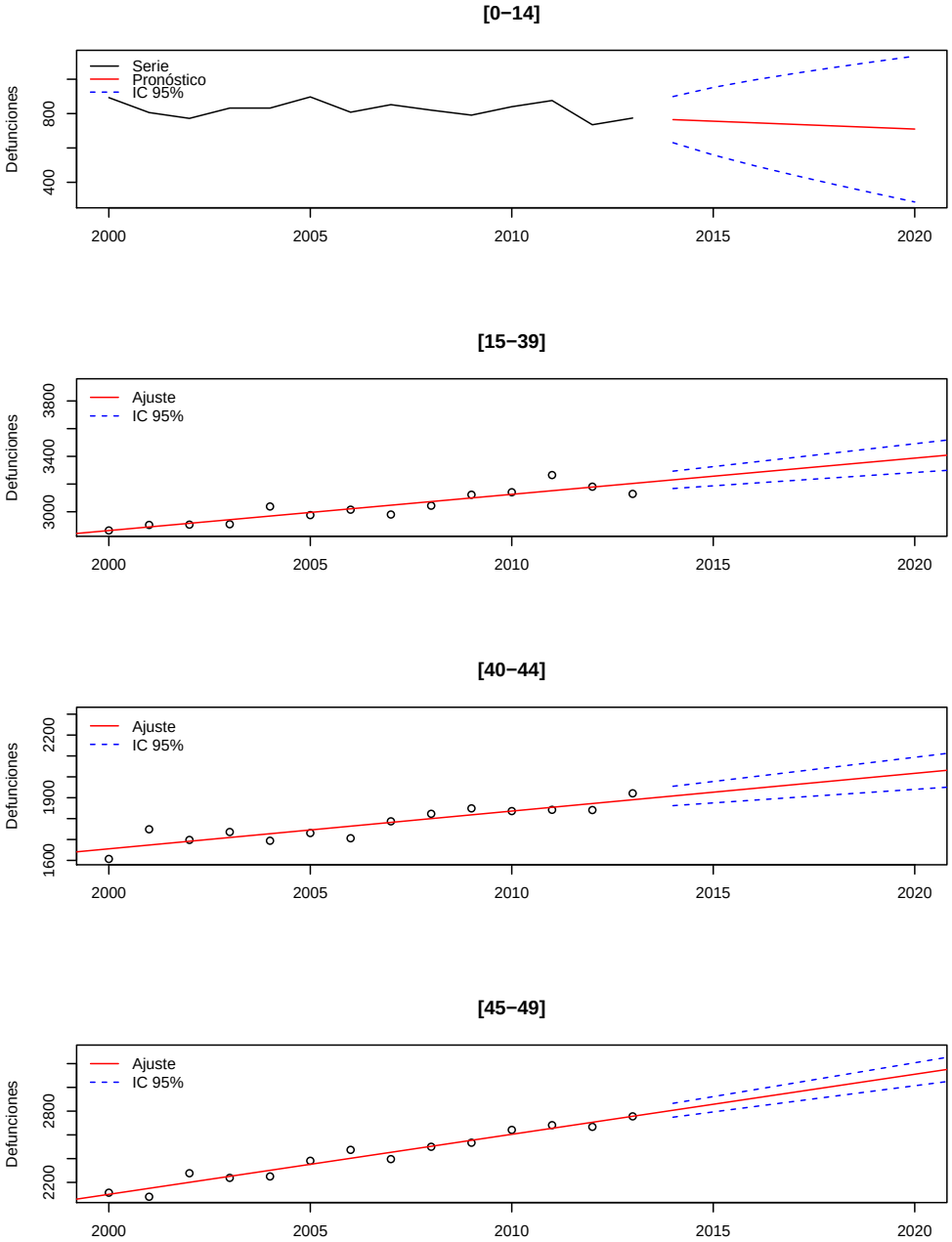


Figura B.11: Defunciones por cáncer por grupos de edad para mujeres

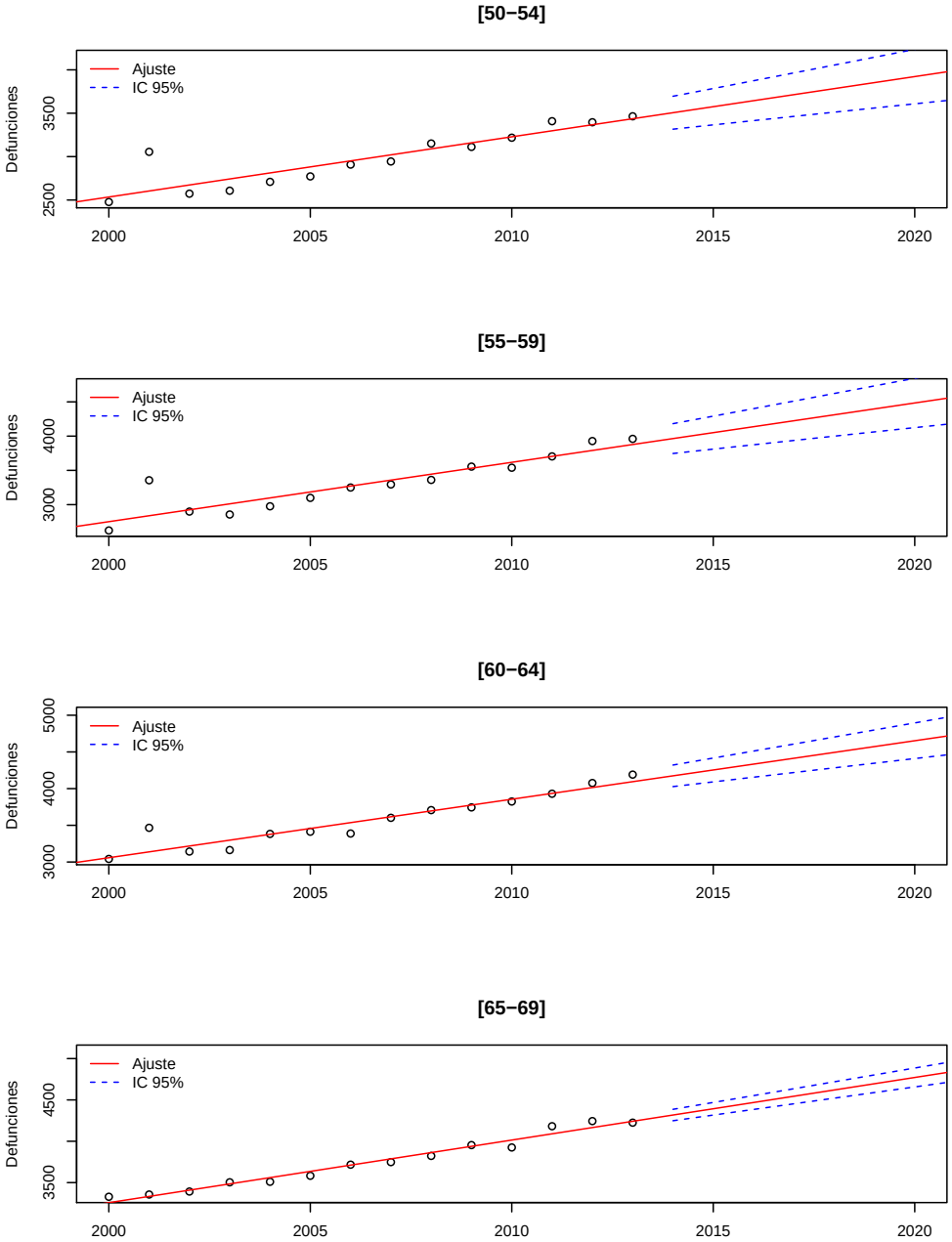


Figura B.12: Defunciones por cáncer por grupos de edad para mujeres

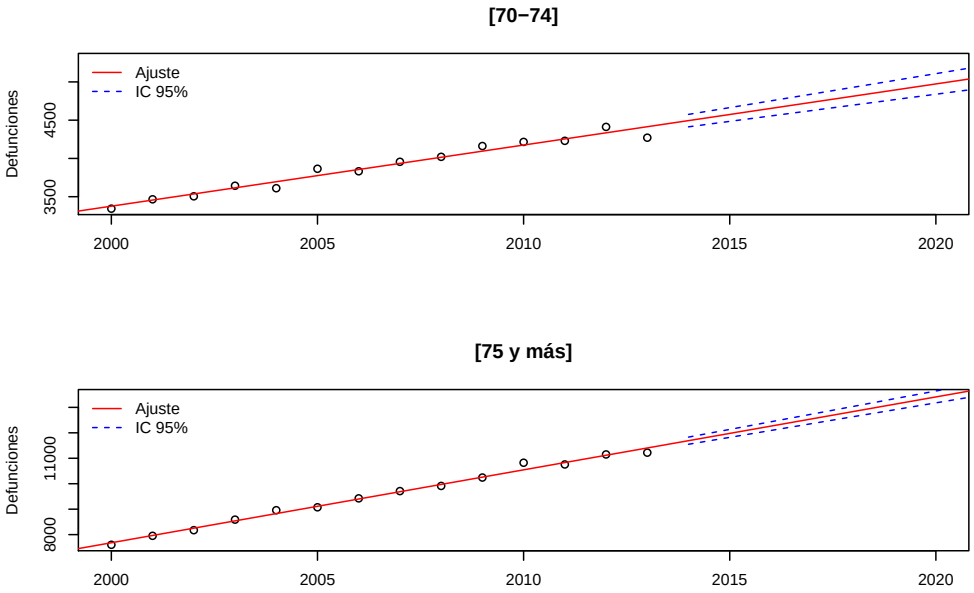


Figura B.13: Porcentaje de la PEA por grupos de edad para hombres

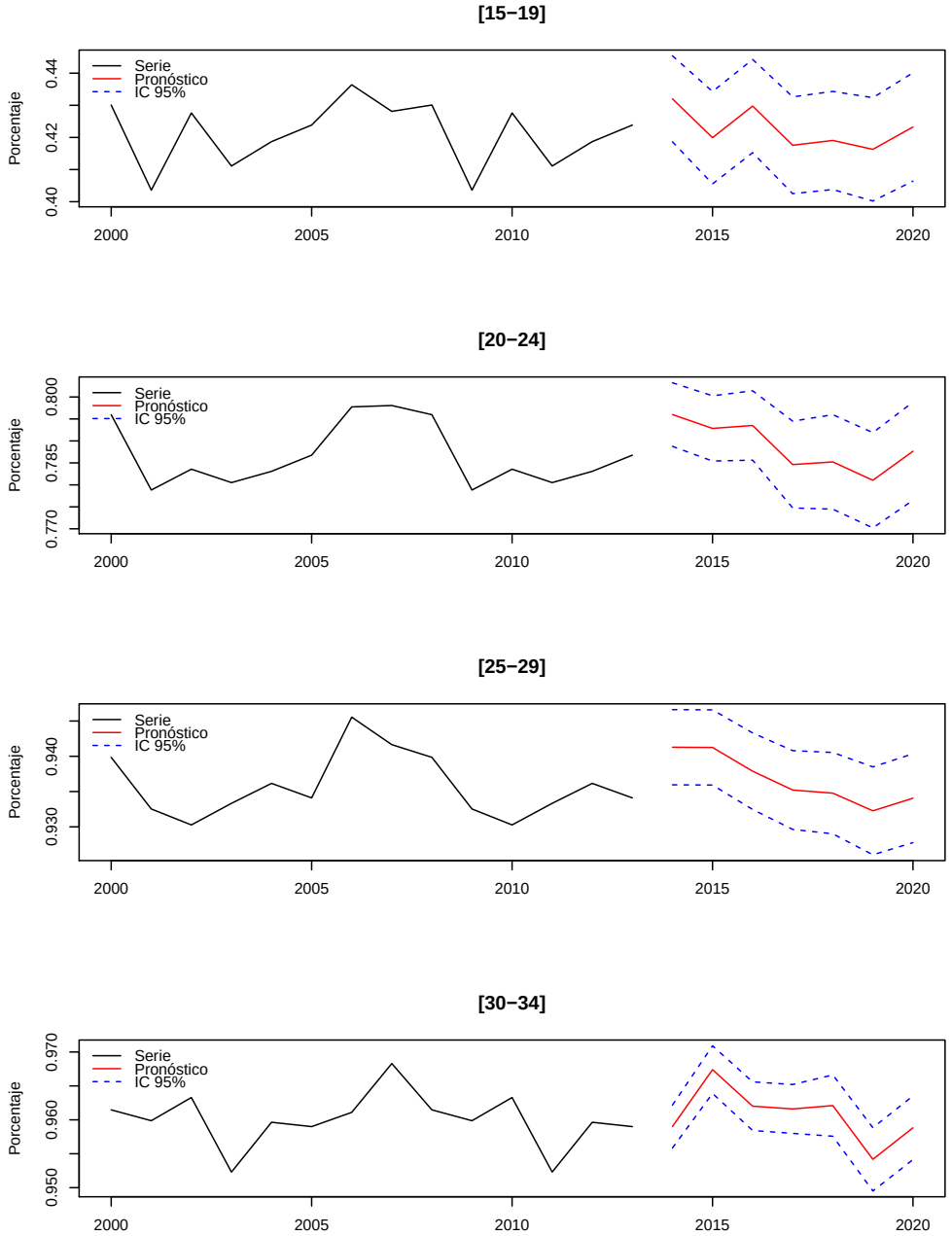


Figura B.14: Porcentaje de la PEA por grupos de edad para hombres

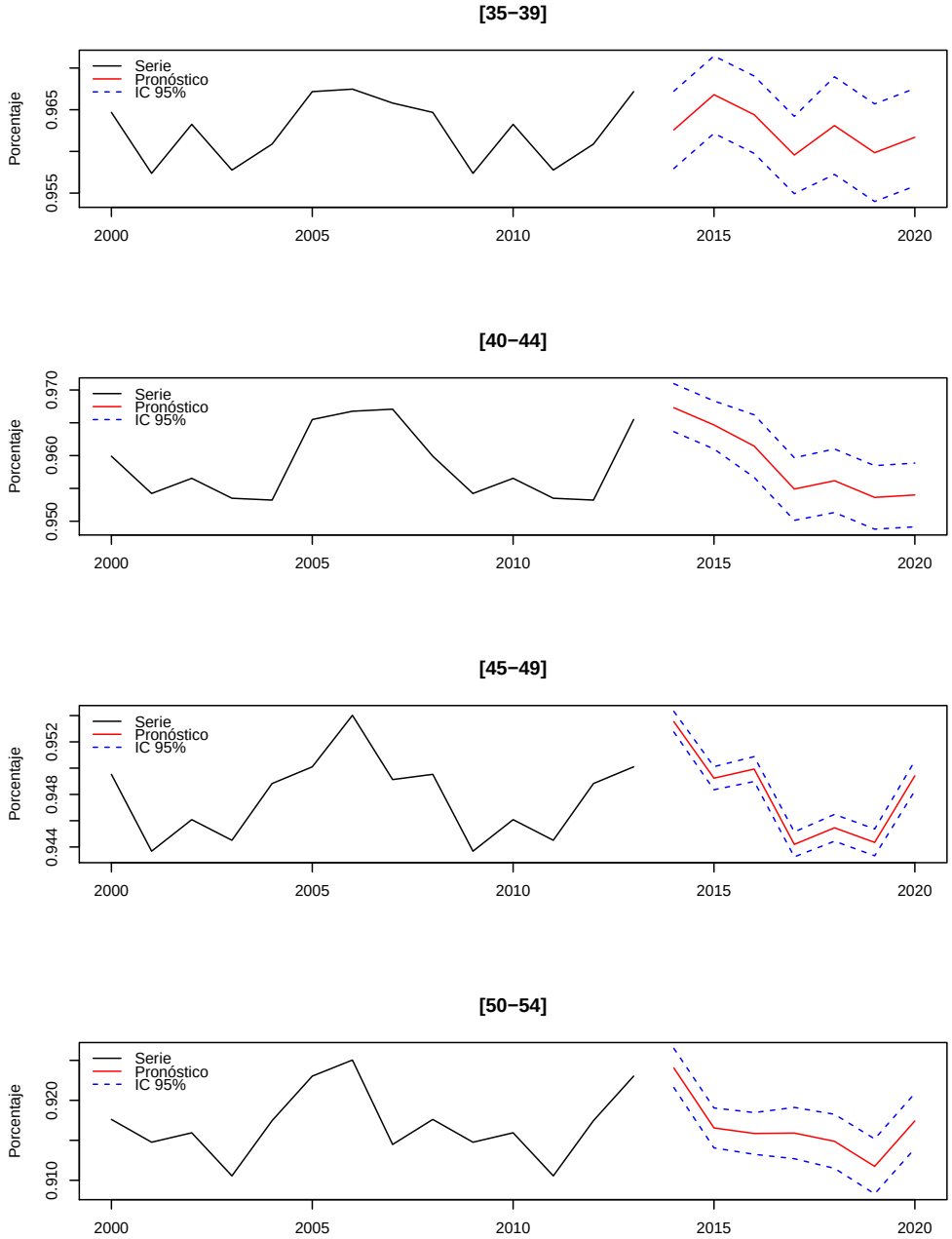


Figura B.15: Porcentaje de la PEA por grupos de edad para hombres

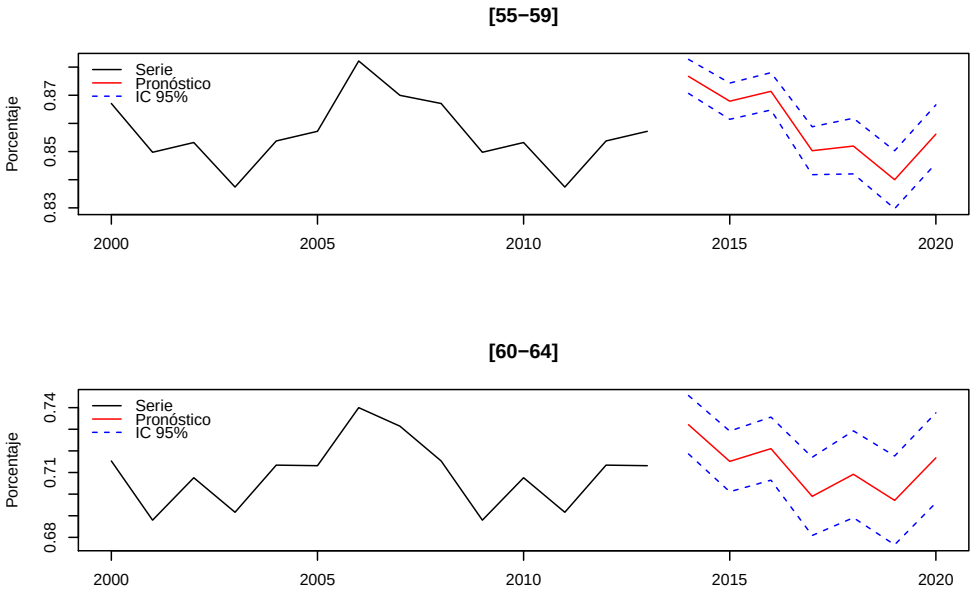


Figura B.16: Porcentaje de la PEA por grupos de edad para mujeres

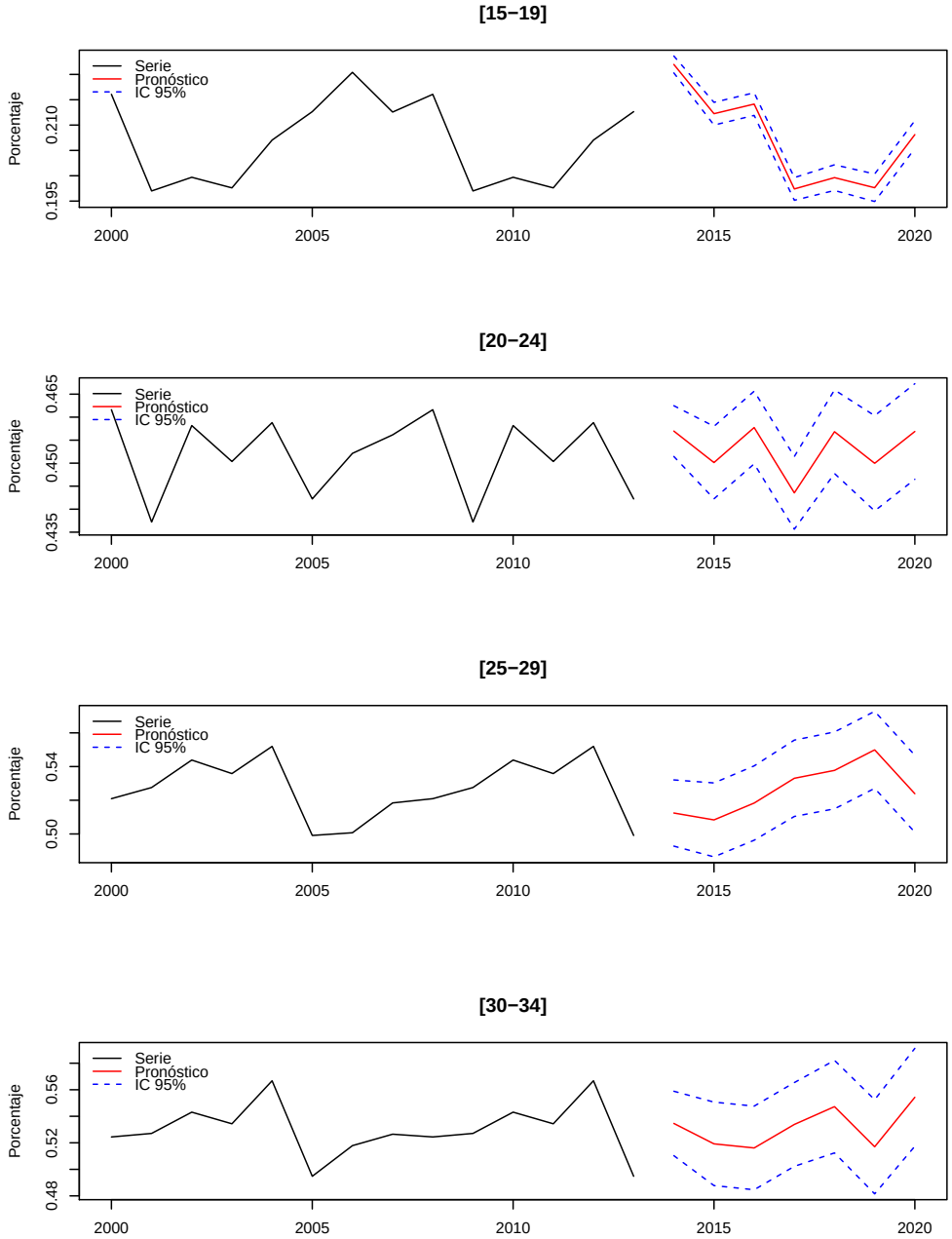


Figura B.17: Porcentaje de la PEA por grupos de edad para mujeres

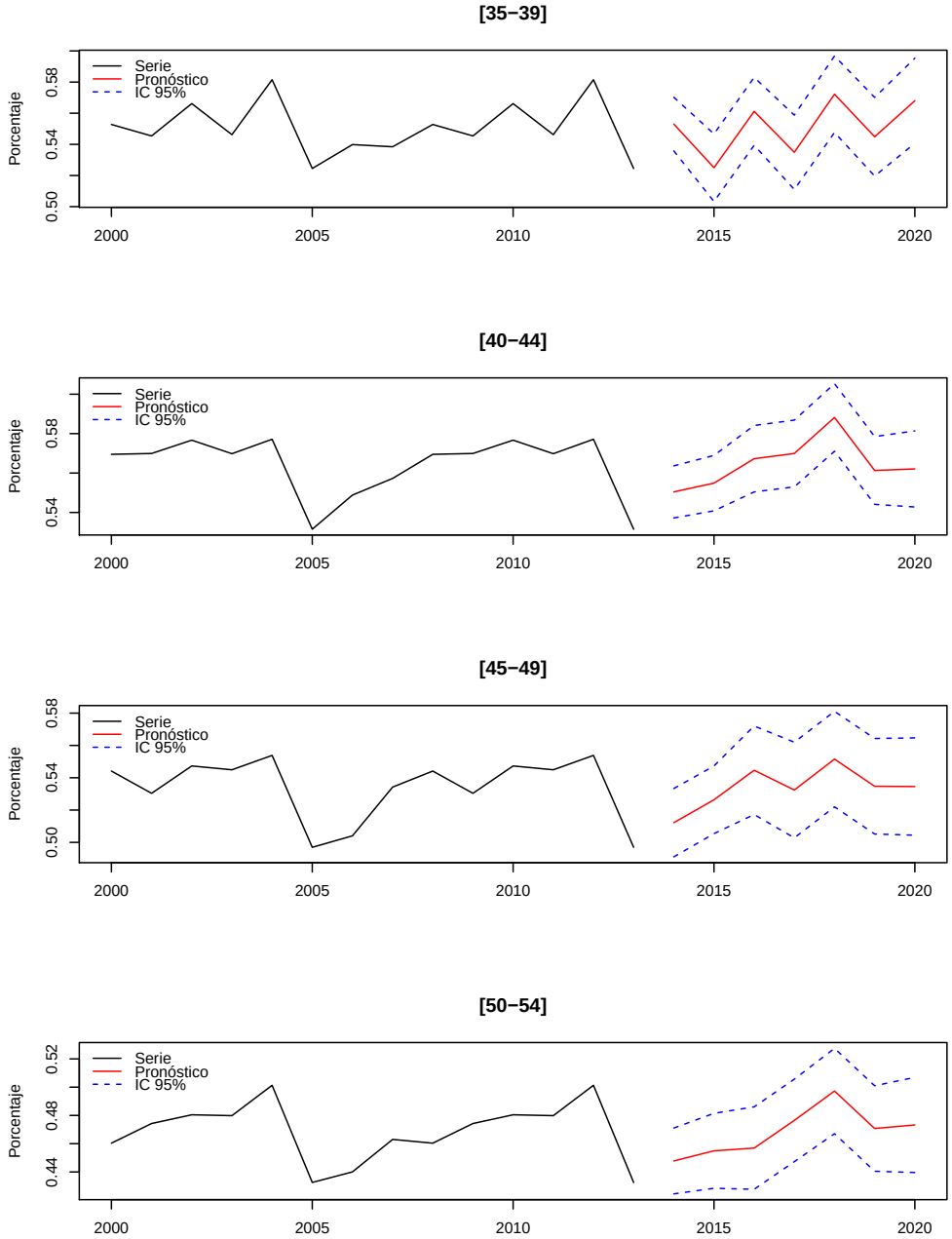


Figura B.18: Porcentaje de la PEA por grupos de edad para mujeres

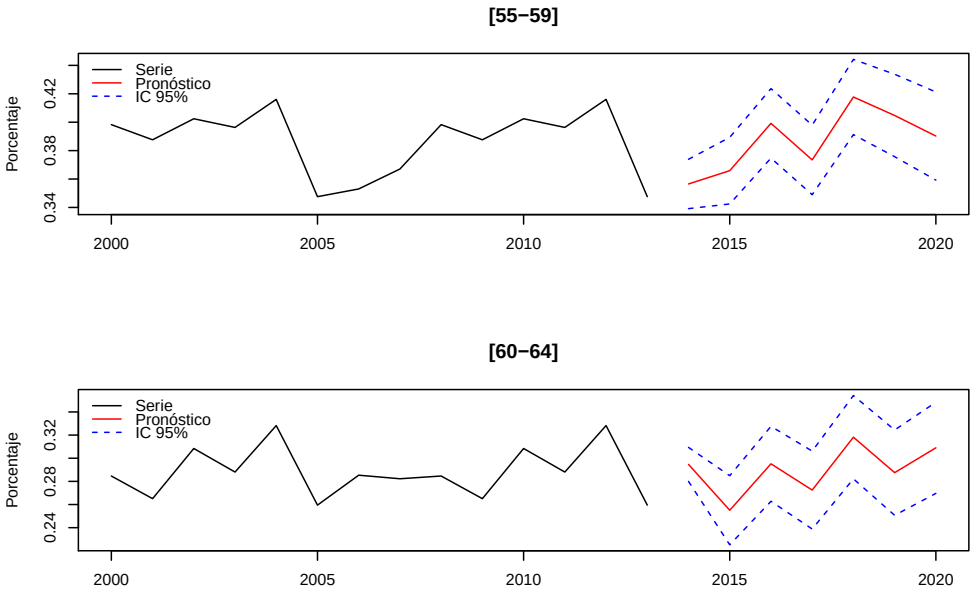


Figura B.19: Porcentaje de la PEAO por grupos de edad para hombres

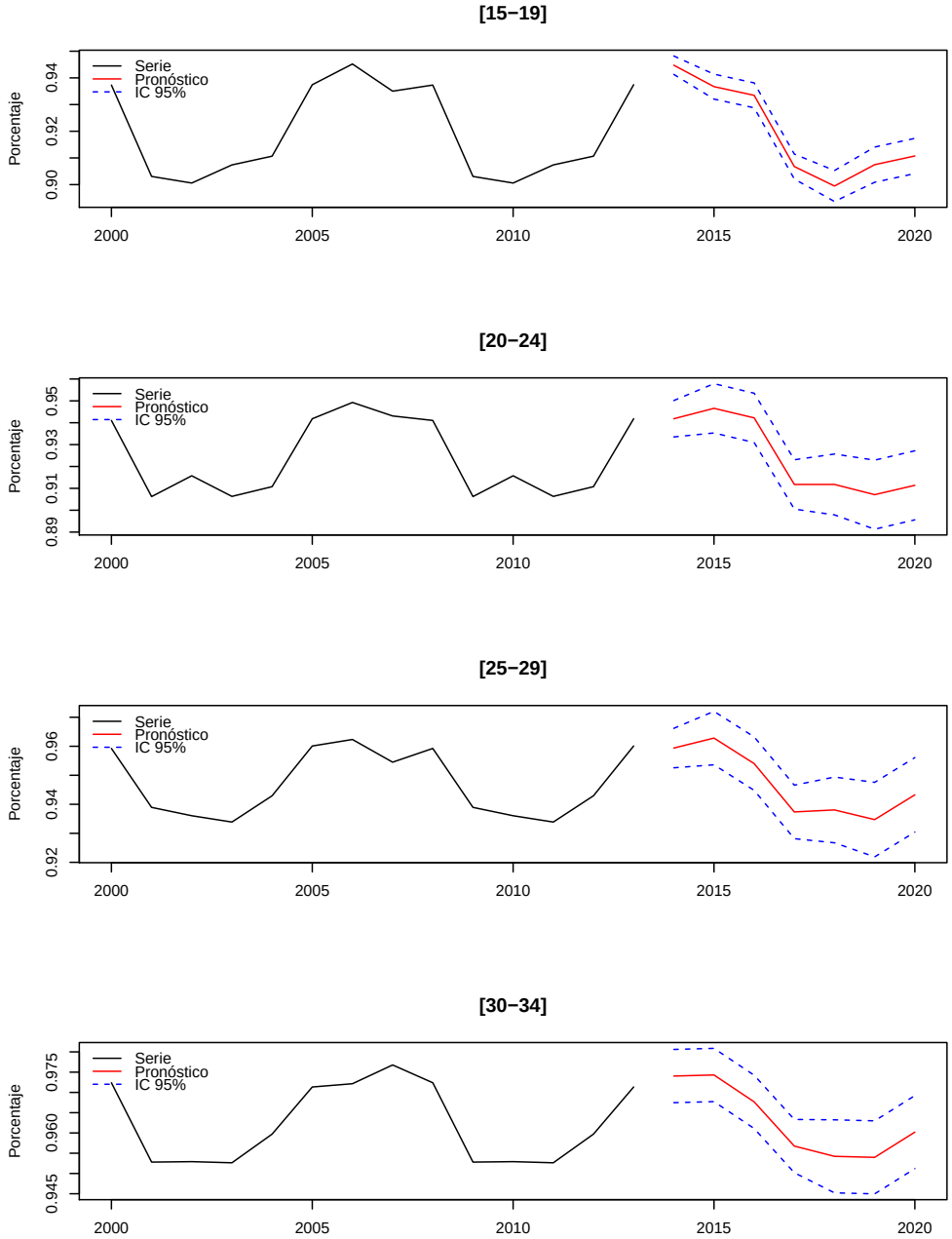


Figura B.20: Porcentaje de la PEAO por grupos de edad para hombres

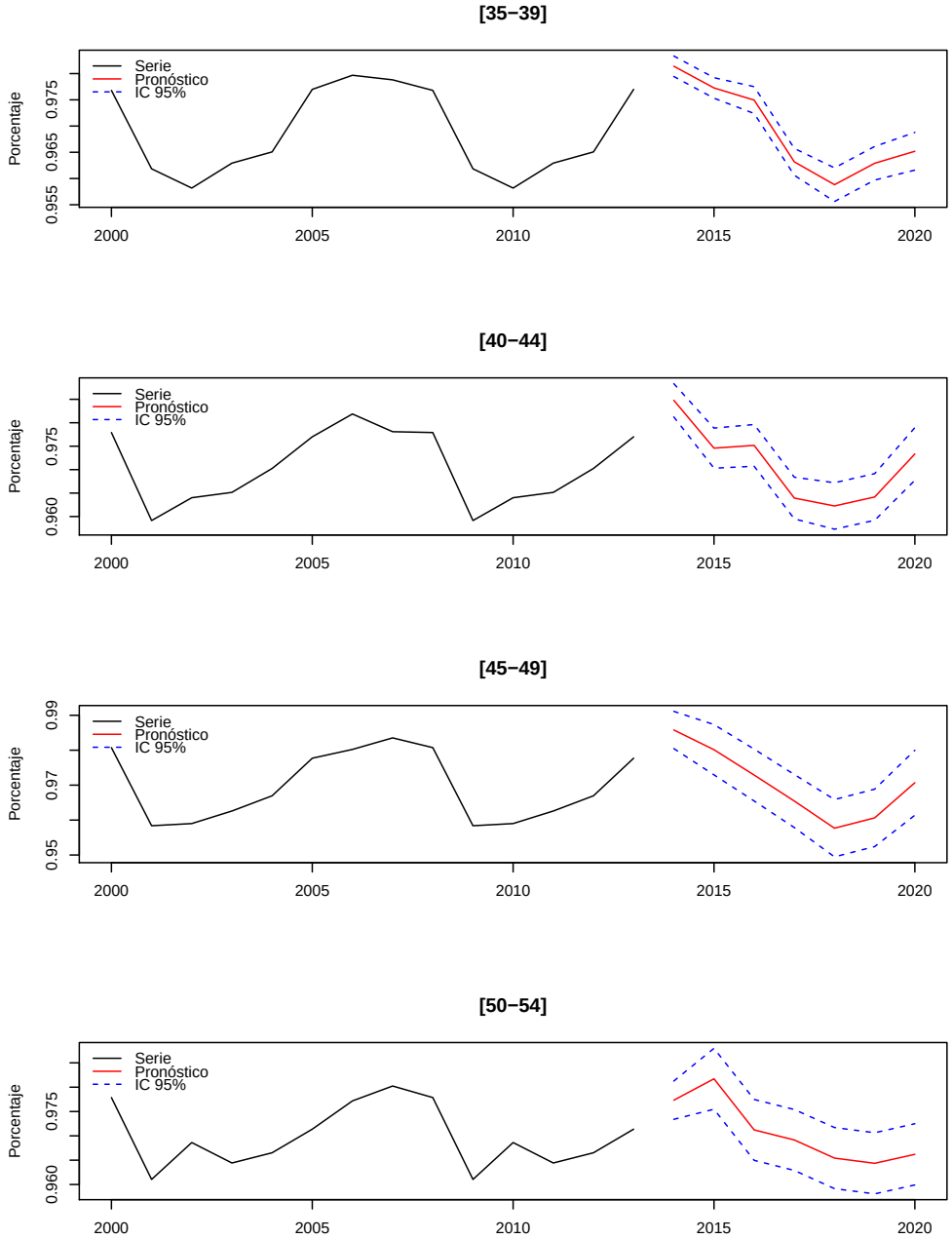


Figura B.21: Porcentaje de la PEAO por grupos de edad para hombres

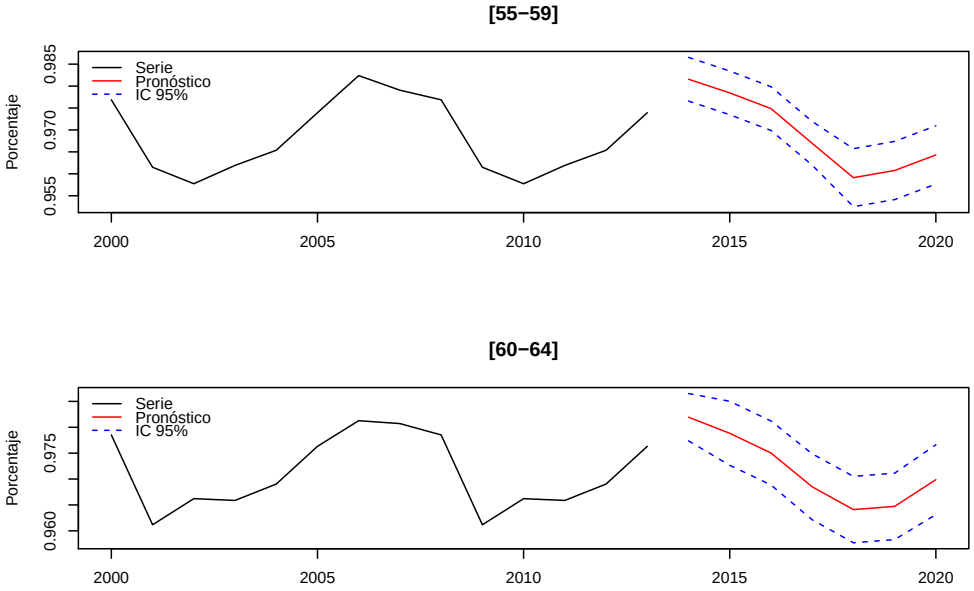


Figura B.22: Porcentaje de la PEAO por grupos de edad para mujeres

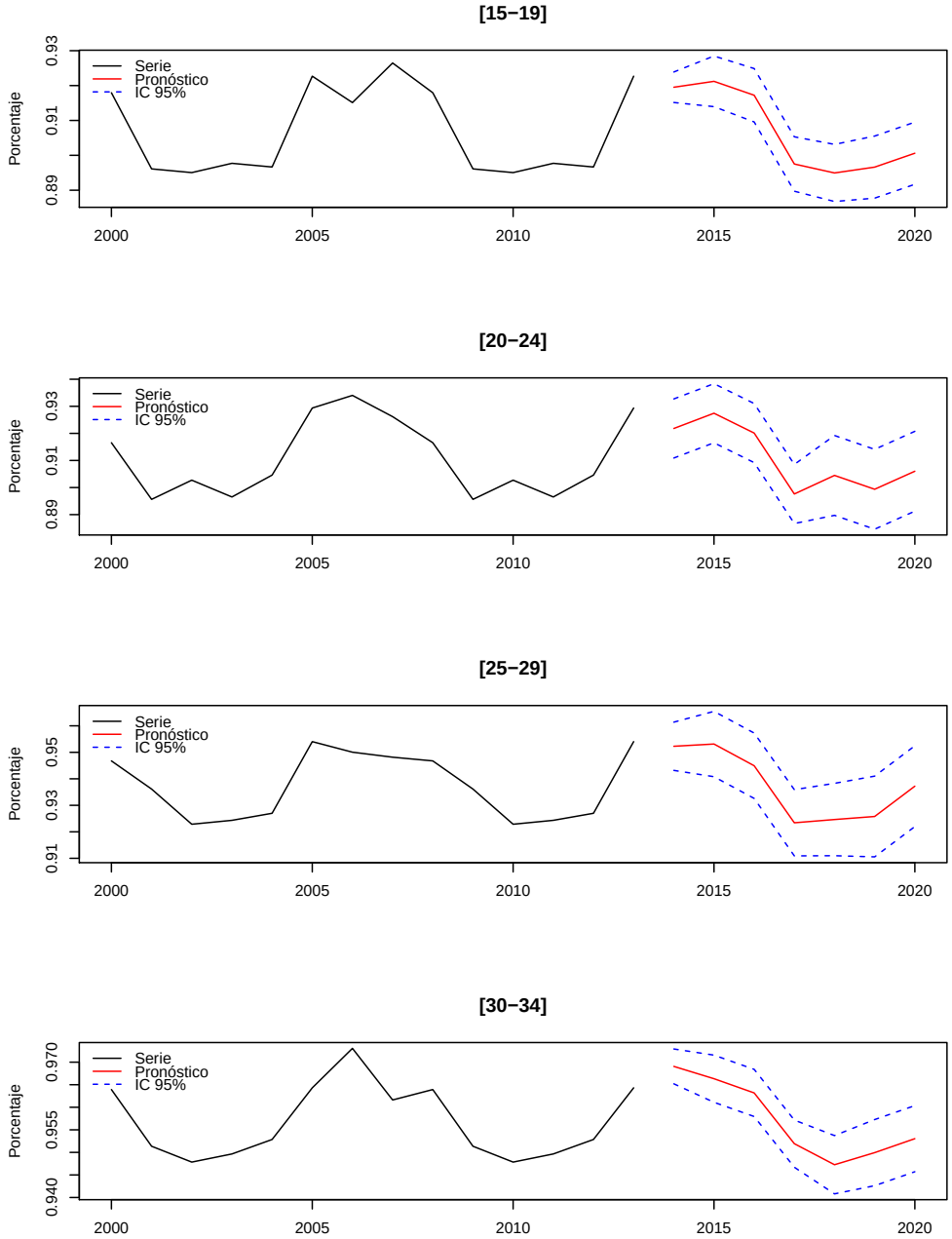


Figura B.23: Porcentaje de la PEAO por grupos de edad para mujeres

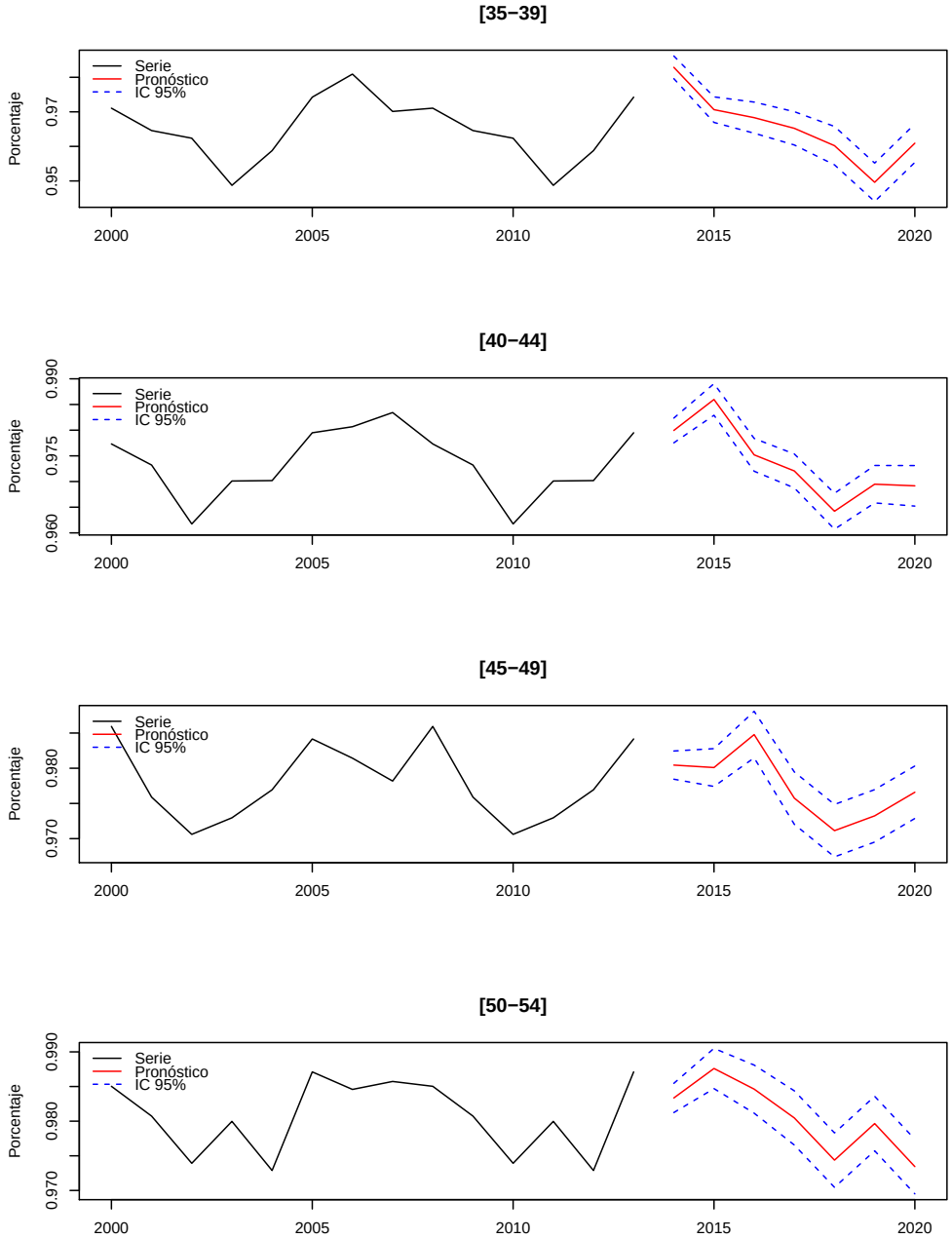


Figura B.24: Porcentaje de la PEAO por grupos de edad para mujeres

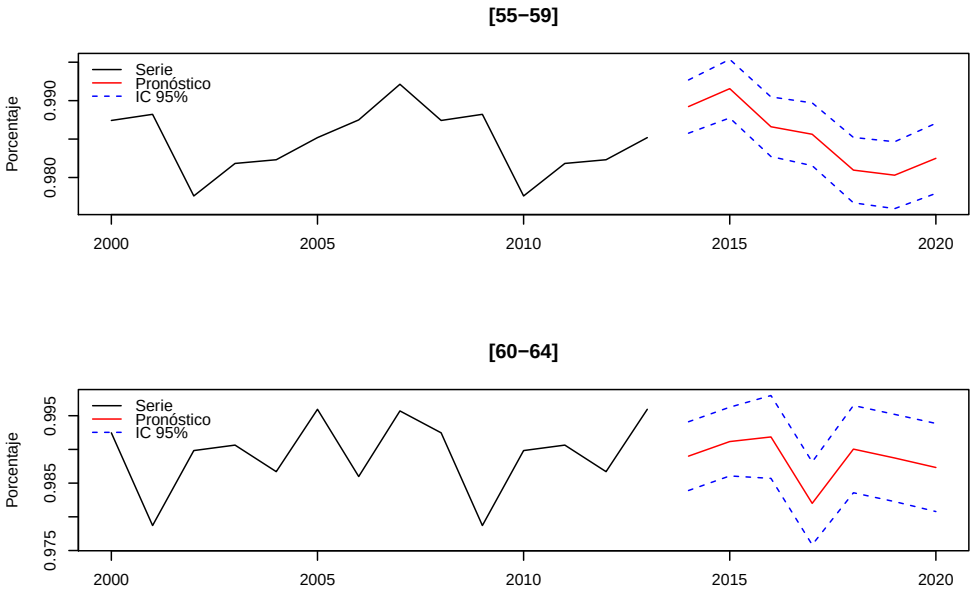


Figura B.25: Salario Base de Cotización correspondiente al Seguro de Enfermedades y Maternidad y al Seguro de Invalidez y Vida

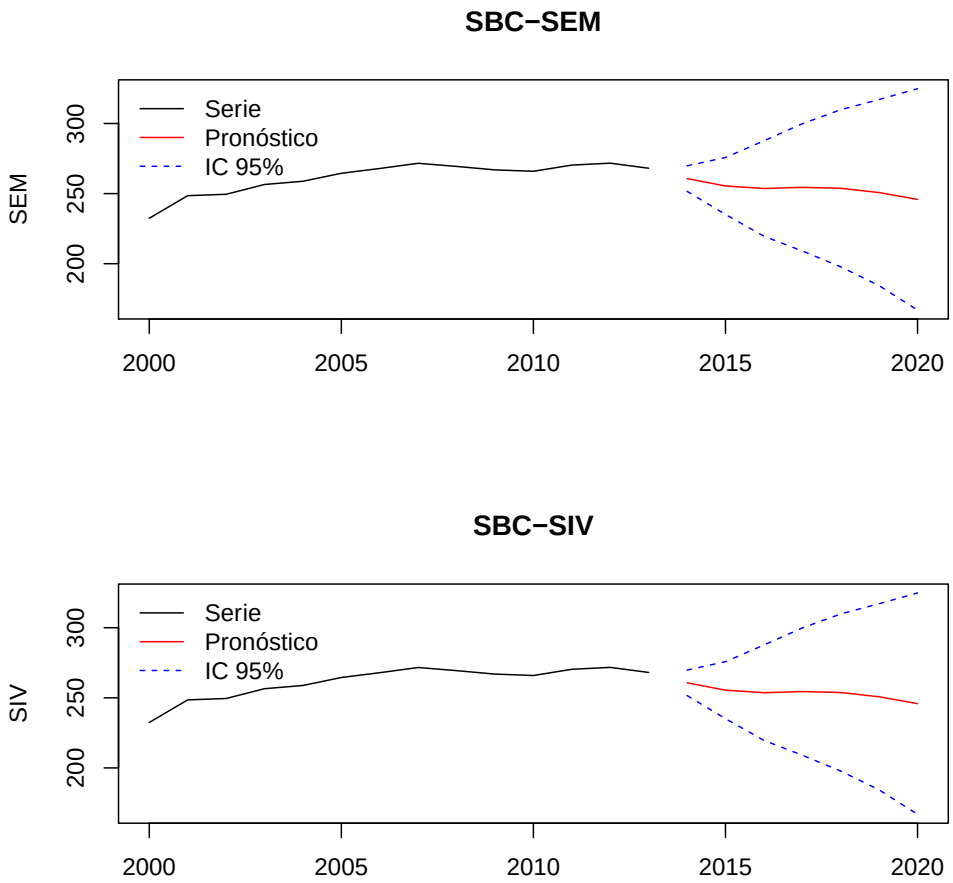


Figura B.26: Ingreso anual de la ENOE por grupos de edad para hombres

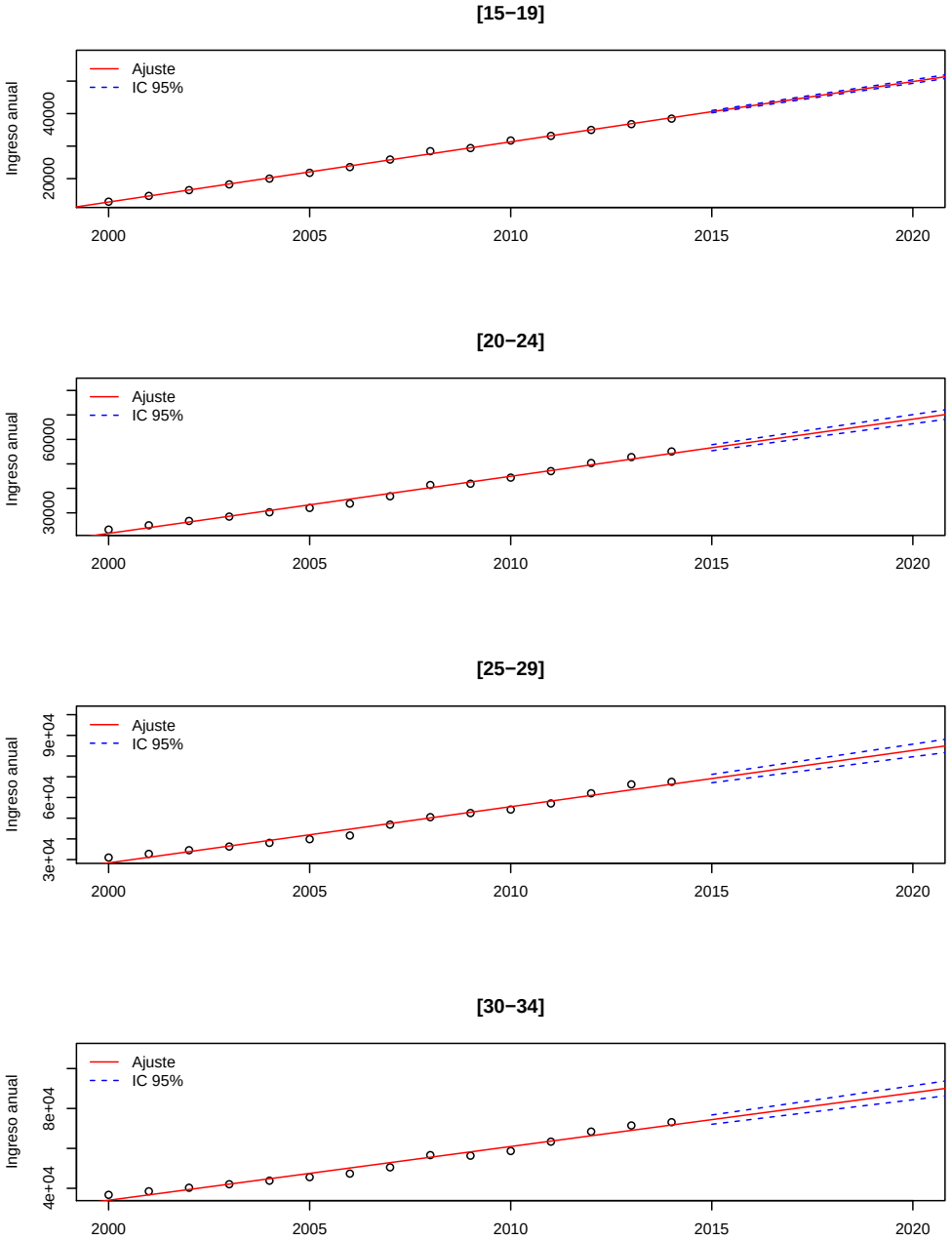


Figura B.27: Ingreso anual de la ENOE por grupos de edad para hombres

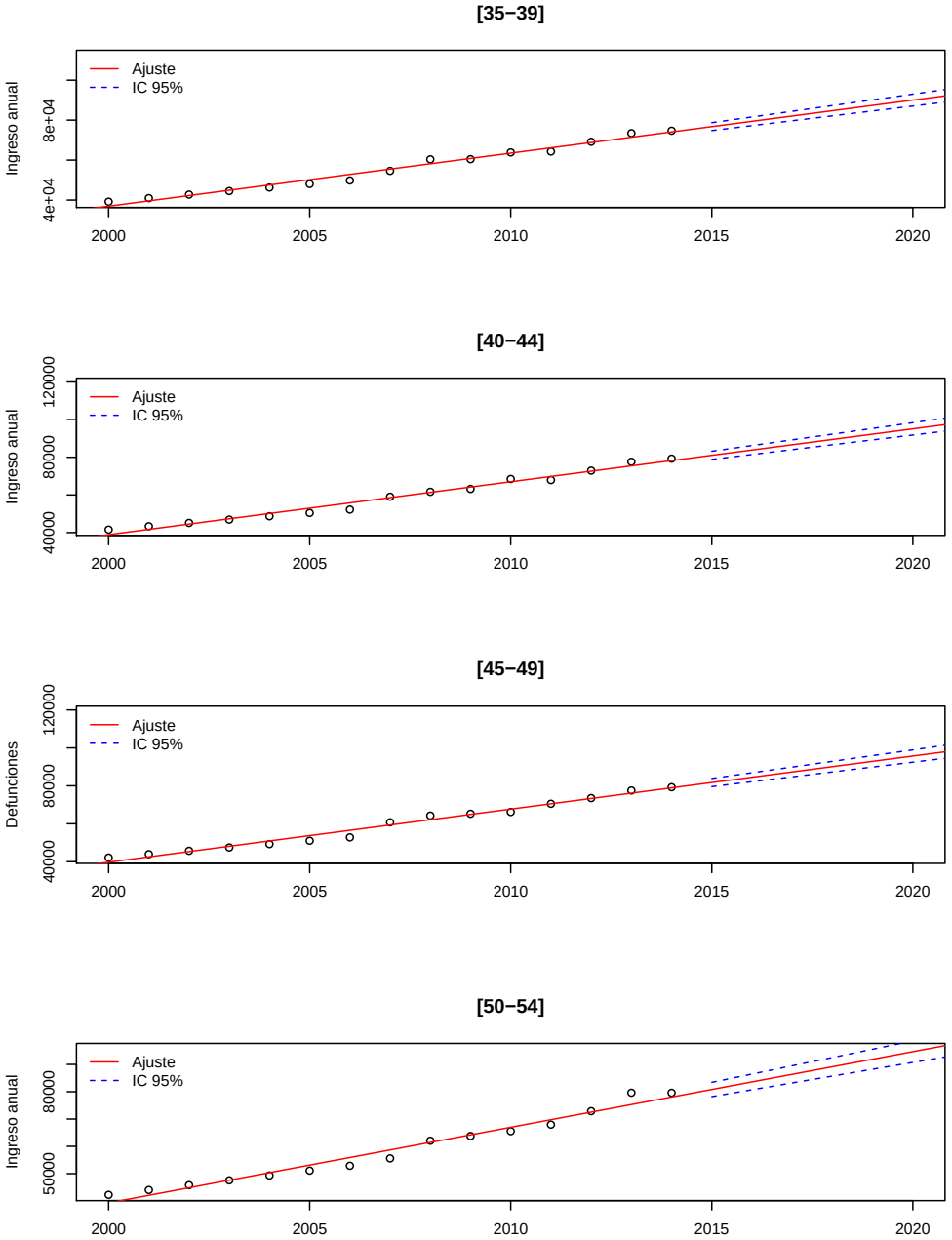


Figura B.28: Ingreso anual de la ENOE por grupos de edad para hombres

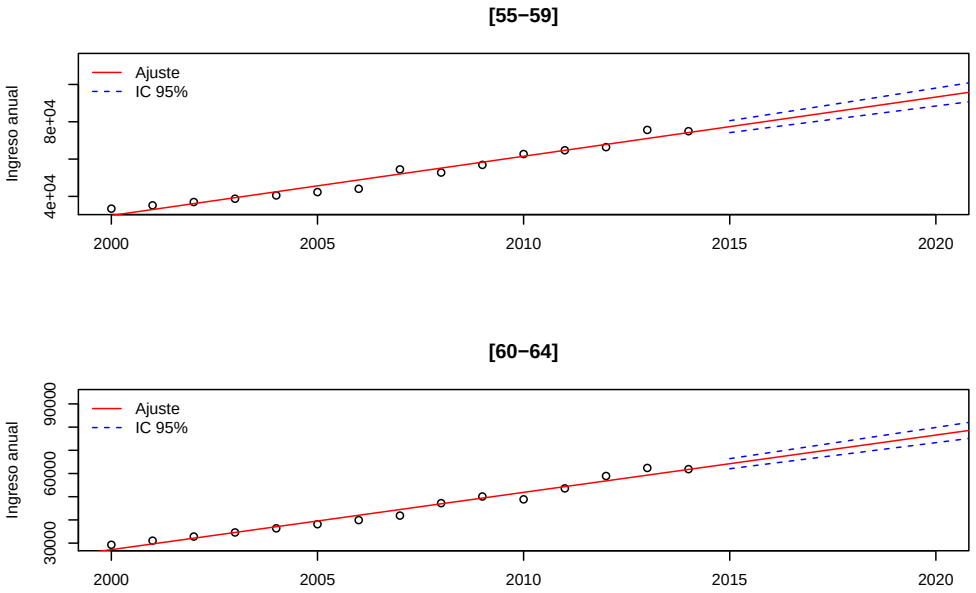


Figura B.29: Ingreso anual de la ENOE por grupos de edad para mujeres

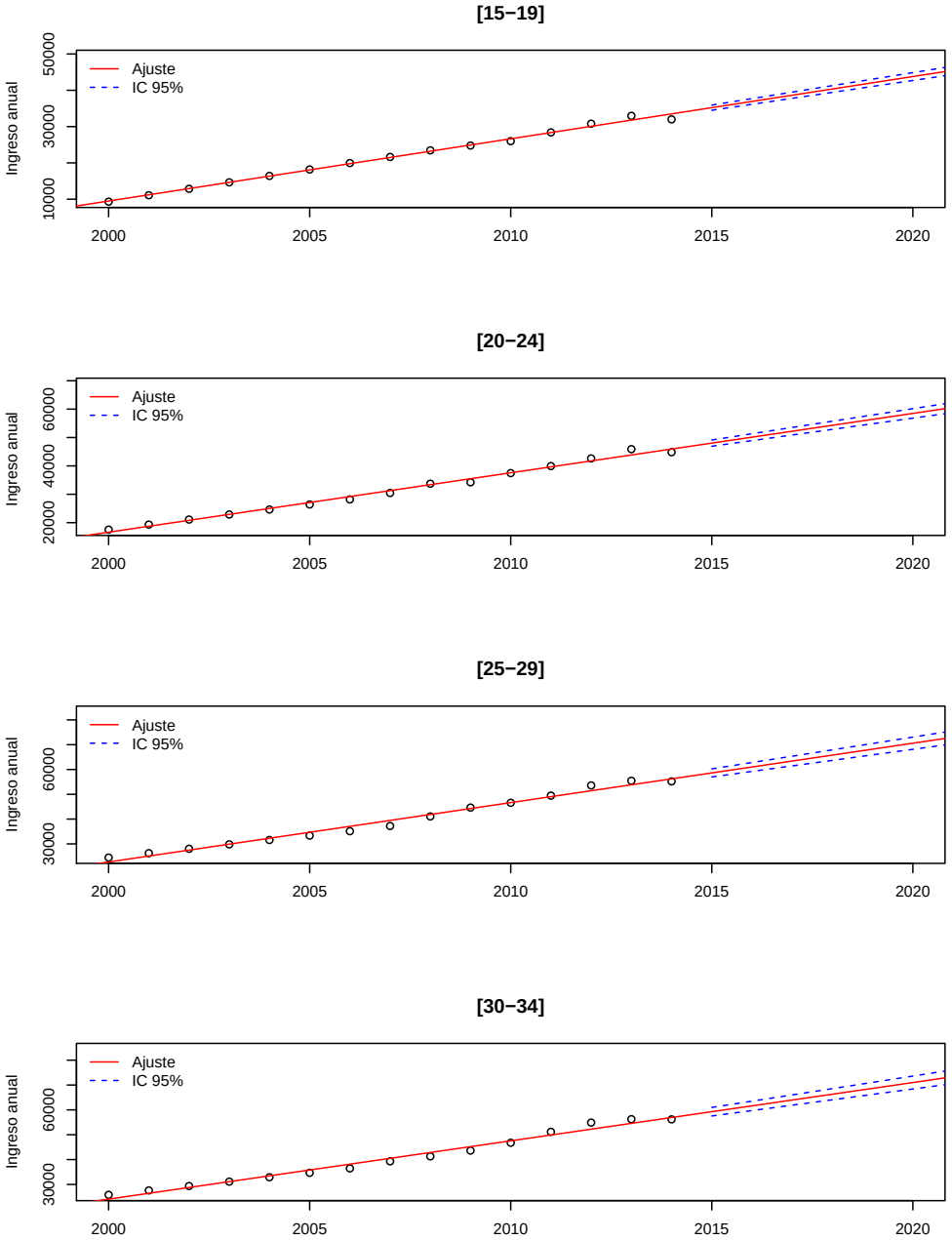


Figura B.30: Ingreso anual de la ENOE por grupos de edad para mujeres

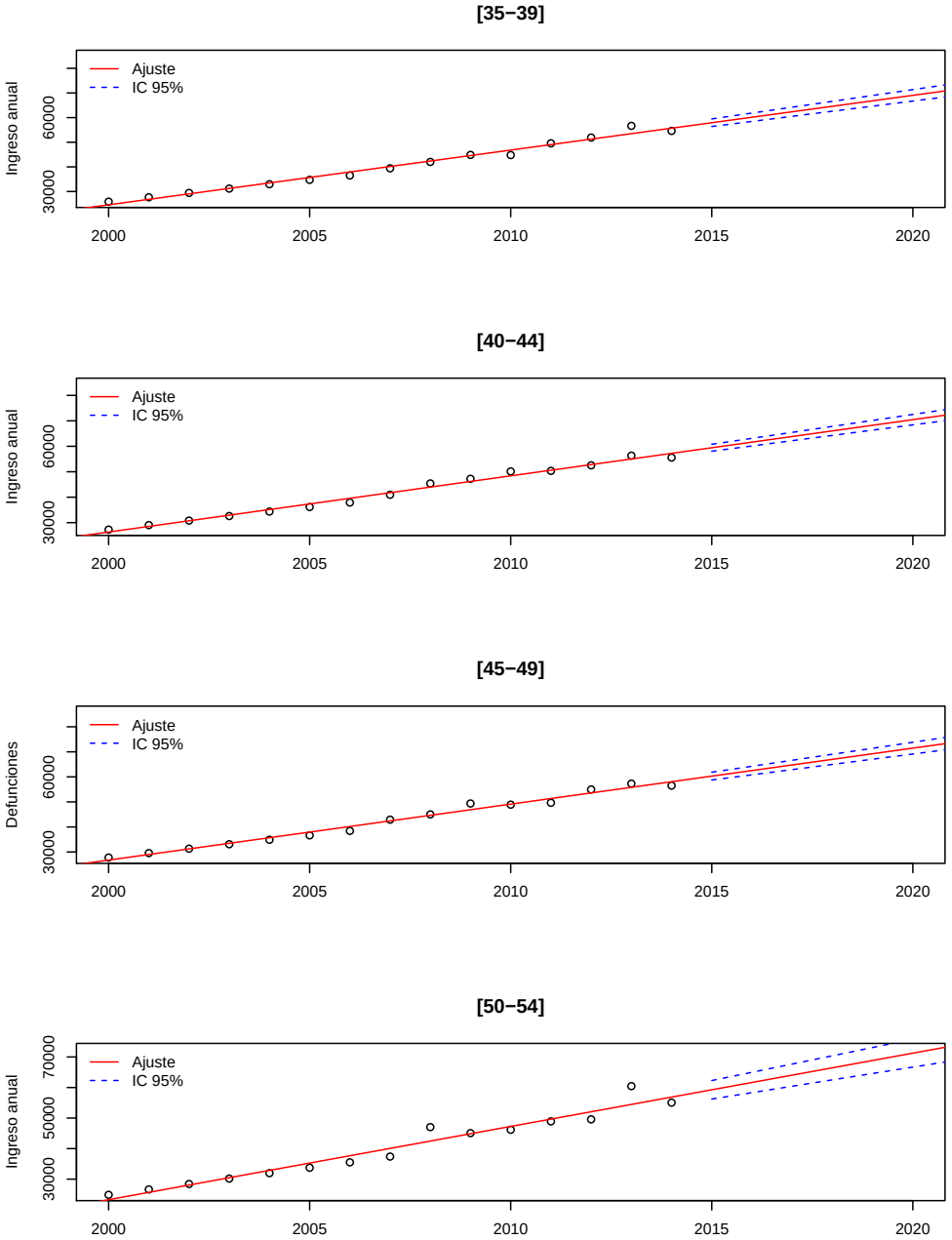


Figura B.31: Ingreso anual de la ENOE por grupos de edad para mujeres

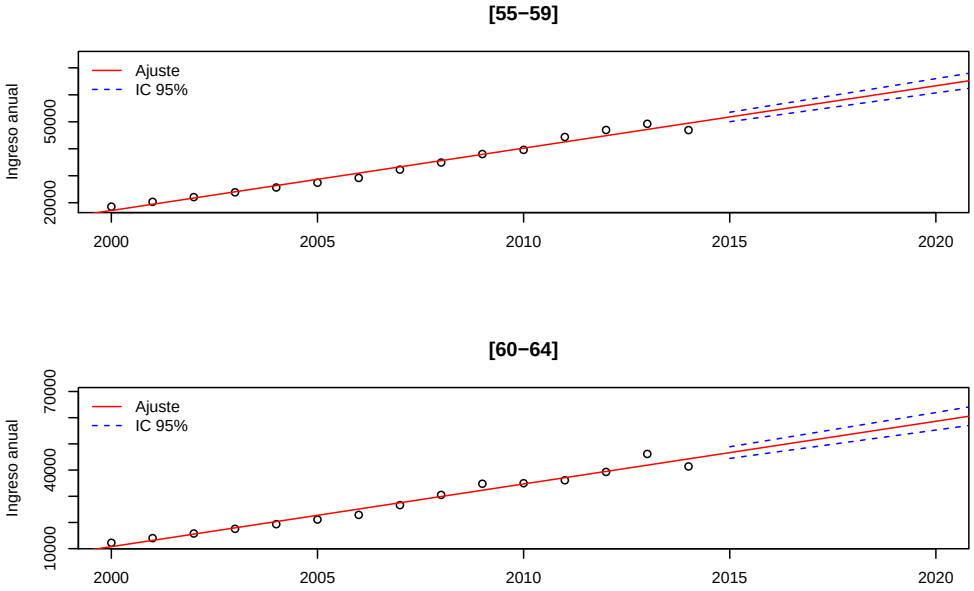


Figura B.32: Porcentaje de la población asalariada divida por sexo.