

---

# MASTER EN BIOESTADÍSTICA Y BIOINFORMÁTICA



Universitat Oberta  
de Catalunya



UNIVERSITAT<sub>DE</sub>  
BARCELONA

## ANÁLISIS ESTADÍSTICO DE EVENTOS ASOCIADOS A VARIABLES DE TIEMPO EN R: MODELO DE SUPERVIVENCIA EN PACIENTES CON CARCINOMA DE CÉLULAS RENALES

Trabajo Final de Master

**Autor:**

Sebastian Alejandro Zapata Acevedo

**Área de trabajo:**

Bioestadística, Aplicación práctica en modelos de enfermedades en el ámbito biomédico

**Tutor:**

Fernando Carmona Berraquero

**Fecha entrega:**

2 de enero de 2018

---

---

*“Entia non sunt multiplicanda sine necessitate”*  
*Principio de la parsimonia (Navaja de Occam)*  
*Siglo XIV (Monje Francisco de Occam)*

---

---

## Ficha de trabajo

Título del trabajo:	Análisis estadístico de eventos asociados a variables de tiempo en R: modelo de supervivencia en pacientes con carcinoma de células renales
Nombre del autor:	Sebastian Alejandro Zapata Acevedo
Nombre del consultor/a:	Fernando Carmona Berraquero
Nombre del PRA:	Carles Ventura Royo
Fecha de entrega (mm/aaaa):	01/2018
Titulación::	Master Bioinformática y Bioestadística
Área del Trabajo Final:	Trabajo Final de Master
Idioma del trabajo:	Español
Palabras clave	Survival data, Applied survival analysis, Survival analysis with R, Survival analysis, Carcinoma, Cáncer, survival analysis models

---

---

## Resumen

El análisis de supervivencia es una de las herramientas estadísticas que permite examinar y modelar el tiempo que tarda un evento en ocurrir. En esta tesis se han estudiado métodos para abordar el análisis de supervivencia enfocados en datos del ámbito biomédico. La base de datos sobre la que se trabajó, contiene un estudio de pacientes con cáncer de células renales, un tipo de enfermedad que representa en la actualidad del 2% al 3% de todos los tipos de cáncer y cuya influencia máxima se da en los países occidentales. Se evaluó en primer lugar la función de supervivencia de para las dos situaciones de eventos (muerte y recaída) asociados a variables de tiempo. La situación de exitus por enfermedad presentó una media de supervivencia de 8 años con un 49% de probabilidad de supervivencia. Para la situación de recaída los casos no llegan al 50% de pacientes que presentaron este evento. Consiguientemente se estimaron las funciones de supervivencia para las variables explicativas del estudio y se realizó un enfoque sobre el género, con el fin de relacionar con otros estudios bibliográficos. Con las estimaciones de la función de supervivencia, observamos que el género de los hombres en un rango de edad entre 50-70 años resultan ser los grupos más vulnerables a padecer esta enfermedad. En segundo lugar se analizó mediante un modelo semiparamétrico de Cox la influencia de las variables respuesta sobre las diferentes situaciones. Se creó un modelo significativo donde las variables dicotomías de quimioterapia y radioterapia resultaron ser las que más influencia presentan sobre los modelos. Por último se evaluó un modelo paramétrico con finalidad de evaluar si nuestros datos podrían explicarse a partir de una distribución conocida y evaluar la diferencia entre los métodos paramétricos y los no paramétricos. Sin embargo, la transformación de los datos para evaluar estos modelos resulta sesgar el estudio y por lo tanto no permite evaluar efectivamente las situaciones, donde los modelos semiparamétrico resultan ser más efectivos.

---

---

## Abstract

The survival analysis is one of the statistical tools that allows us to examine and model the time for the occurrence of an event. In this thesis, methods have been studied to approach survival analysis focused on data from the biomedical field. The database on which we work contains a study about patients with renal cell cancer, a type of disease that currently represents 2% to 3% of all types of cancer and whose maximum influence is in the Western countries. The survival function was first evaluated for the two situations of events (death and relapse) associated with time variables. The situation of death due to illness presented an average survival of 8 years with a 49% probability of survival. For the relapse situation, cases do not reach 50% of patients who presented this event. Consequently, the survival functions were estimated for the explanatory variables of the study and a focus on gender was carried out, in order to relate it with other bibliographic studies. With the estimates of the survival function, we observed that the gender of men in an age range between 50-70 years are the most vulnerable groups to suffer of this disease. On the other hand, the influence of the response variables on the different situations was analyzed by means of a semi-parametric Cox model. A significant model was created where the dichotomies of chemotherapy and radiotherapy turned out to be the most influential on the models. Finally, a parametric model was evaluated in order to evaluate if our data could be explained from a known distribution and evaluate the difference between parametric and non-parametric methods. However, the transformation of the data to evaluate these models results in a bias in the study and, therefore, does not allow an effective evaluation of situations, where semi-parametric models are more effective.

---

---

## Tabla de contenido

---

<b>1.</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
1.1	CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO	2
1.2	OBJETIVOS	3
1.2.1	<i>OBJETIVOS GENERALES</i>	3
1.2.2	<i>OBJETIVOS ESPECÍFICOS</i>	3
1.3	ENFOQUE Y MÉTODO A SEGUIR	3
1.4	PLANIFICACIÓN DEL TRABAJO	4
1.5	BREVE SUMARIO DE PRODUCTOS OBTENIDOS	6
1.6	BREVE DESCRIPCIÓN DE OTROS CAPÍTULOS DE LA MEMORIA	6
<b>2.</b>	<b>DESCRIPCIÓN TEÓRICA DEL PROBLEMA (MARCO TEÓRICO)</b>	<b>7</b>
2.1	ANTECEDENTES	7
2.2	TEORÍA SOBRE EL ANÁLISIS DE SUPERVIVENCIA	8
2.2.1	<i>CONCEPTOS BÁSICOS</i>	9
2.3	PLANTEAMIENTO DEL PROBLEMA	14
2.3.1	<i>DESCRIPCIÓN DE LA BASE DE DATOS</i>	15
2.3.2	<i>CODIFICACIÓN DE VARIABLES</i>	16
2.4	ANÁLISIS ESTADÍSTICO	17
2.4.1	<i>RESUMEN ESTADÍSTICO PARA VARIABLES CONTINUAS Y FACTORES</i>	17
2.4.2	<i>ESTIMACIÓN DE FUNCIÓN DE SUPERVIVENCIA PARA VARIABLES DE TIEMPO</i>	21
2.4.3	<i>COMPARACIÓN DE CURVAS DE SUPERVIVENCIA EN VARIABLES EXPLICATIVAS</i>	29
2.5.1	<i>ESTIMACIÓN DE MODELO DE COX UNIVARIANTE Y MULTIVARIANTE</i>	36
2.5.2	<i>ESTIMACIÓN DEL MODELO DE COX UNIVARIANTE</i>	36
2.5.2	<i>ESTIMACIÓN DE MODELO DE COX MULTIVARIANTE</i>	37
2.4.3	<i>ESTIMACIÓN DE MODELO PARAMÉTRICO</i>	49
2.5	ANÁLISIS DE RESULTADOS	55
<b>3.</b>	<b>CONCLUSIONES</b>	<b>57</b>

---

---

<b>4.</b>	<b>GLOSARIO</b>	<b>58</b>
<b>5.</b>	<b>BIBLIOGRAFÍA</b>	<b>59</b>
<b>6.</b>	<b>ANEXOS</b>	<b>64</b>

---

---

## Figuras

<b>Figura 1.</b> Diagrama de Gantt del proyecto .....	5
<b>Figura 2.</b> Ejemplo de eventos censurados .....	11
<b>Figura 3.</b> Gráfico de distribución de edad .....	19
<b>Figura 4.</b> Función global de supervivencia situación 1 .....	23
<b>Figura 5.</b> Función de supervivencia por sexo situación 1 .....	25
<b>Figura 6.</b> Función global de supervivencia situación 2 .....	27
<b>Figura 7.</b> Función de supervivencia por sexo situación 2 .....	29
<b>Figura 8.</b> Función de supervivencia tipo de tumor situación 1 .....	30
<b>Figura 9.</b> Función de supervivencia tipo de tumor situación 2 .....	30
<b>Figura 10.</b> Función de supervivencia tipo de tumor y sexo situación 1 .....	31
<b>Figura 11.</b> Función de supervivencia tipo de tumor y sexo situación 2 .....	31
<b>Figura 12.</b> Función de supervivencia por edad situación 1 .....	32
<b>Figura 13.</b> Función de supervivencia por edad situación 2 .....	32
<b>Figura 14.</b> Función de supervivencia por edad y sexo situación 1 .....	33
<b>Figura 15.</b> Función de supervivencia por edad y sexo situación 2 .....	33
<b>Figura 16.</b> Función de supervivencia radioterapia y(o) quimioterapia situación 1 .....	34
<b>Figura 17.</b> Función de supervivencia radioterapia y(o) quimioterapia situación 2 .....	34
<b>Figura 18.</b> Función de supervivencia radioterapia y(o) quimioterapia y sexo situación 1 .....	35
<b>Figura 19.</b> Función de supervivencia radioterapia y(o) quimioterapia y sexo situación 2 .....	35
<b>Figura 20.</b> Análisis de residuos y bosques situación 1 .....	43
<b>Figura 21.</b> Análisis de residuos y bosques situación 2 .....	48
<b>Figura 22.</b> Función de supervivencia estimada para Weibull Situación 1 .....	54
<b>Figura 23.</b> Comparación funciones de supervivencia modelos de Cox y Weibull situación 1....	54

---

---

## Tablas

<b>Tabla 1.</b> Variables de estudio .....	15
<b>Tabla 2.</b> Resumen descriptivo variables de estudio .....	17
<b>Tabla 3.</b> Relación variables tipo de tumor y sexo .....	18
<b>Tabla 4.</b> Distribución de edad según sexo .....	20
<b>Tabla 5.</b> Comparación modelos paramétricos situación 1 .....	53
<b>Tabla 6.</b> Comparación modelos paramétricos situación 2 .....	53

---

---

## 1. Introducción

Nos encontramos en una época donde todo el entorno físico y biológico, genera una cantidad inmensa de datos pertenecientes a varias disciplinas de interés para la sociedad. Estos se encuentran organizados de diferente manera en función de la utilidad que se les pretende dar. Ante esta gran disponibilidad de datos se enfrentan las herramientas informáticas, las cuales deben ser lo suficientemente potentes para tratarlos y analizarlos de forma que se pueda extraer la máxima información posible, mediante la cual se pueda comprender cada vez mejor su naturaleza y variación con el objetivo de poder predecir su comportamiento.

En este caso, nos enfrentamos ante datos clínicos de pacientes que padecen *carcinoma de células renales*. Estos datos se encuentran en su mayoría muy organizados y los márgenes de error al analizarlos deben ser cada vez más pequeños, ya que en medio se encuentra la comprensión de una de las enfermedades quizás más desafiantes hasta la época, y de mayor interés para la investigación.

Según el último informe "World cancer report 2014" realizado por la *IARC (International Agency for Research on Cancer)*, el cáncer se sitúa como una de las principales causas de morbilidad y mortalidad en el mundo. Aproximadamente 10 millones de personas son diagnosticadas de cáncer anualmente y más de 6 millones mueren por esta enfermedad todos los años, cifra que va en aumento. Se prevé que el número de casos de cáncer aumente a 22 millones en las siguientes dos décadas (Stewart, BWKP 2014). Ante esta situación las herramientas que nos ayuden a comprender y predecir el desarrollo y evolución de esta enfermedad se convierten en un soporte fundamental. Aquí es donde aparece el análisis de supervivencia. Esta herramienta ayuda a la predicción de la enfermedad identificando los factores pronósticos (sexo, estadio, tratamiento) relacionados con el evento (recidiva o fallecimiento) y, según las estadísticas sobre tiempo de supervivencia de datos registrados con anterioridad, ayuda a realizar un acercamiento descriptivo y evolutivo de los pacientes que padecen esta enfermedad (Gómez, Guadalupe 2004).

El análisis estadístico sobre el estudio de enfermedades se ha potenciado en los últimos años gracias a las herramientas informáticas disponibles, capaces de sacar el máximo aprovechamiento de los datos. Estas herramientas, además, permiten un mejor análisis,

---

facilitando así la comprensión y posterior comparación con otras investigaciones, ayudando a avanzar progresivamente en el conocimiento de las enfermedades.

## **1.1 Contexto y justificación del Trabajo**

Es cierto que muchos campos de estudio se esmeran por intentar predecir el momento de ocurrencia de algún evento de su interés, es por eso que incorporan el concepto dinámico del tiempo como variable principal mediante la cual se intenta describir el momento en el que sucede el evento «time-to-event» (Le, Chap T 1997). El presente trabajo toma como punto de partida, el tiempo, que es una de las unidades dimensionales más importante en las disciplinas académicas y de investigación. Este tiempo lo asociamos con eventos o sucesos, en nuestro caso es la muerte o recaída del paciente, que junto a los datos recopilados en las variables de seguimiento y apoyados en análisis estadísticos, permitirá la estimación de las diferentes funciones de supervivencia y evaluación de modelos ajustados sobre la base de datos. Estos métodos nos ayudarán a comprender, analizar y predecir los tiempos de duración de una determinada situación que culmina con la ocurrencia de un evento y su relación con otros factores clínicos demográficos de seguimiento, los cuales pueden catalizar la ocurrencia del mismo.

El carcinoma de células renales (CCR) supone el 3% de las enfermedades malignas del adulto (Torrelles, M, Garc et al., 2006). Debido al avance en las diferentes tecnologías de detección, el número de tumores diagnosticados tempranamente ha aumentado en estadios iniciales I y II, permitiendo tratamientos más conservadores y que aportan buenos resultados sobre la supervivencia de los pacientes. Sin embargo, el CCR se sigue caracterizando por presentar un curso clínico variable e impredecible (Ljungberg, B et al., 2010), creando así, la necesidad de estudios y técnicas analíticas que ayuden al entendimiento de esta patología.

El presente trabajo se centra en datos de pacientes diagnosticados con CCR, dentro del ámbito biomédico. La disciplina mediante la cual analizamos estos datos es la bioestadística, la técnica de investigación es la conocida como análisis de supervivencia y la herramienta de computación que utilizamos para estos análisis, es R, un lenguaje rico en funciones de análisis de supervivencia y de gran apoyo para las investigaciones en este campo (Views, R 2017).

---

## **1.2 Objetivos**

### **1.2.1 Objetivos generales**

- Estudiar eventos asociados a variables de tiempo en R (modelos de supervivencia).
- Realizar una aplicación práctica de análisis de supervivencia en una base de datos real de pacientes con carcinoma de células renales.

### **1.2.2 Objetivos específicos**

- Realizar una revisión bibliográfica enfocada en el desarrollo de análisis de supervivencia con R.
- Estimar la función de supervivencia en relación con las variables categóricas y cuantitativas de la base de datos de pacientes con CCR.
- Identificar los diferentes modelos teóricos de supervivencia.
- Ajustar un modelo considerando las variables de seguimiento sobre la base de datos de pacientes CCR.
- Realizar una evaluación comparativa de fortalezas y debilidades de cada modelo (paramétricos vs no paramétricos).

## **1.3 Enfoque y método a seguir**

La metodología utilizada para la realización del presente trabajo se fundamenta en dos partes:

- La investigación y recopilación de material bibliográfico enfocada a modelos de datos de supervivencia o longitudinales.
- La selección de investigaciones donde son aplicados los análisis de supervivencia usando el programa *R*, con el fin de desarrollar y aplicar nuestros propios modelos y funciones de supervivencia sobre base de datos clínicos de pacientes CCR.

---

## 1.4 Planificación del trabajo

A continuación, se desglosa la estrategia a seguir en diferentes tareas:

- 1) Búsqueda bibliográfica utilizando las palabras clave (Análisis de Supervivencia, Survival Analysis, Survival Analysis with R). Una vez terminada la revisión realizamos las siguientes tareas:
  - ➔ Compilación de información.
  - ➔ Criba de la información relevante.
  - ➔ Análisis de modelos y función de supervivencia con paquetes de R.
- 2) Limpieza y transformación de variables sobre la base de datos de pacientes con carcinoma en células renales.
- 3) Planteamiento de modelos y función de supervivencia sobre nuestra base de datos.
- 4) Análisis estadístico de los resultados obtenidos y comparación con los modelos bibliográficos.
- 5) Redacción de conclusiones.
- 6) Elaboración de la memoria final con los resultados de la investigación.
- 7) Defensa de la investigación ante el tribunal del la UOC-UB.

Se desarrolla la estrategia más coherente con el fin de alcanzar los objetivos planteados, tomando como pilar básico los fundamentos bibliográficos y la posterior aplicación de los mismos para el desarrollo de los modelos de supervivencia sobre la base de datos en la cual se realiza este trabajo.

La planificación del trabajo con el diagrama de Gantt, se ha realizado mediante la aplicación *Gantter* de *Google Drive*. Se ha seguido el listado de tareas enumerado en el epígrafe anterior, teniendo en cuenta, asimismo, los hitos de la programación del aula en relación a la entrega de los PECs.

# Diagrama de Gantt para el proyecto TFM

ID	Nombre	Type	% Completado	Inicio	Fin	Predecesoras
0	Diagrama de Gantt para el proyecto TFM	Ganther	88%	20/09/2017	22/01/2018	
1	PEC0	Ganther	100%	20/09/2017	02/10/2017	
2	PEC1	Ganther	100%	03/10/2017	16/10/2017	
3	Búsqueda bibliográfica	Ganther	100%	20/09/2017	10/11/2017	
4	PEC2	Ganther	100%	17/10/2017	20/11/2017	
5	Limpieza y transformación de variables de la base de datos	Ganther	100%	13/11/2017	17/11/2017	3
6	PEC3	Ganther	100%	21/11/2017	18/12/2017	
7	Creación y análisis estadístico de modelos y funciones de supervivencia para el dataset	Ganther	100%	21/11/2017	18/12/2017	5,3
8	PEC4	Ganther	100%	19/12/2017	10/01/2018	
9	Redacción de la memoria	Ganther	100%	19/12/2017	02/01/2018	7
10	PEC5	Ganther	0%	11/01/2018	22/01/2018	
11	Elaboración de la presentación	Ganther	0%	03/01/2018	10/01/2018	9
12	Presentación del TFM	Ganther	0%	11/01/2018	22/01/2018	11

ID	Risk	Response	Cause	Probability	Severity	Priority
1	Error en los modelos creados con la base de datos			Probable	Moderate	Urgent action
2	Errores con los paquetes de R			Occasional	Moderate	Urgent action
3	Problemas con el ordenador			Improbable	Moderate	Action

ID	Nombre del Recurso	Email	Tipo	Coste	Calendario Bas
1	Bibliografía referente a Análisis de Supervivencia		Trabajo	€0/h	Estándar
2	Bibliografía referente a paquetes "R" para análisis de supervivencia		Trabajo	€0/h	Estándar
3	Bases de datos pacientes con Carcinoma		Trabajo	€0/h	Estándar

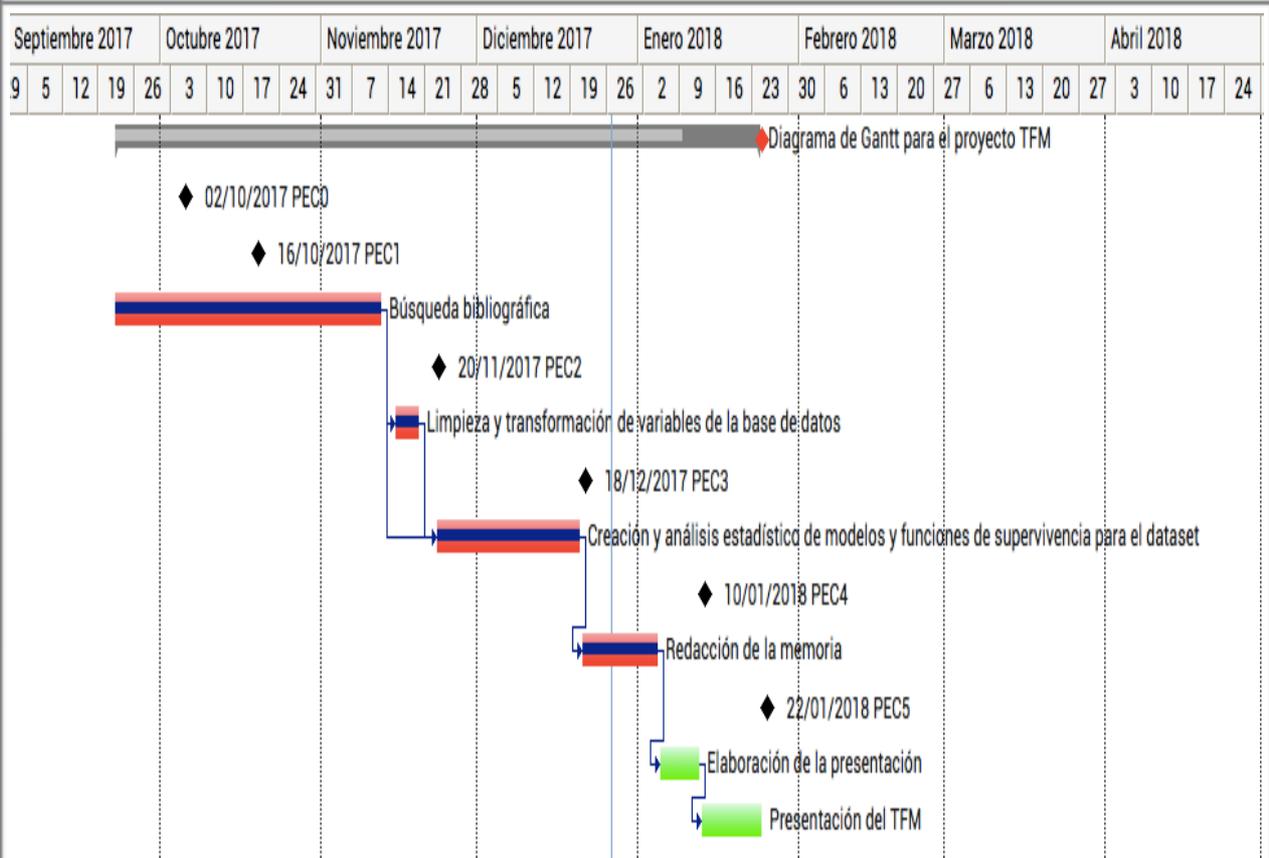


Figura 1: Diagrama de Gantt

---

## **1.5 Breve resumen de productos obtenidos**

De acuerdo con los objetivos trazados, los productos obtenidos en la realización del presente proyecto son:

- Una base bibliográfica sólida que recopila documentos para el desarrollo de análisis de supervivencia en R.
- El análisis de función de supervivencia para las diferentes variables explicativas de la base de datos con un enfoque en la variable sexo.
- El análisis estadístico mediante método no paramétricos y paramétricos de supervivencia utilizando en R sobre la base de datos de pacientes CCR.
- Evaluación y análisis comparativo de la metodología más apropiada en relación con la base de datos de pacientes con CCR.

## **1.6 Breve descripción de otros capítulos de la memoria**

Desarrollada la introducción y planteados los objetivos, a continuación, en los siguientes capítulos, se recopilan los fundamentos teóricos y posterior análisis estadístico de este trabajo. El Capítulo 2 contiene el marco teórico - práctico del proyecto y se encuentra dividido, a su vez, en varios epígrafes :

- Antecedentes: describe la vista previa de la problemática del análisis de supervivencia desde una perspectiva histórica, intentando resaltar la importancia que ha tenido esta técnica en los últimos años.
- Teoría sobre el análisis de datos: desarrolla fundamentos de la teoría del análisis de supervivencia y sus diferentes aplicaciones en la investigación biomédica.
- Planteamiento del problema: describe los fundamentos teóricos a aplicar para el desarrollo del análisis estadístico de la base de datos.
- Descripción de la base de datos: desglosa cada una de las variables y transformaciones que se deben aplicar sobre la base de datos original con el objetivo de seleccionar y preparar las variables a utilizar en el análisis estadístico.
- Análisis estadístico: recopila los resultados y discusión de los valores obtenidos utilizando R.

---

El Capítulo 3 describe las conclusiones procedentes del análisis estadístico de supervivencia sobre la base de datos real de pacientes con carcinoma de células renales.

Los capítulos restantes (Glosario, Bibliografía, Anexos), constituyen parte formal de la memoria y detallan términos utilizados, documentos bibliográficos soporte de la investigación e información extra que se considera relevante para la investigación, pero no se encuentra detallada dentro de la estructura del proyecto.

## **2. Descripción teórica del problema (Marco teórico)**

### **2.1 Antecedentes**

El análisis de supervivencia es un tema en auge en la actualidad, potenciado en los últimos años gracias al desarrollo de tecnologías informáticas capaces tanto de almacenar como analizar complejos volúmenes de datos (Views, R 2017). El análisis de supervivencia es una de las ramificaciones más importantes de la estadística, su origen data del siglo XVII (Views, R 2017) gracias a la construcción de tablas de mortalidad por *Edmon Halley* (astrónomo) y su evolución estuvo fuertemente influenciada por la ingeniería en la segunda guerra mundial con el objetivo de analizar la duración y fiabilidad de los diferentes elementos que forman las máquinas (Gómez, Guadalupe 2004). Los modelos de análisis de supervivencia tienen aplicaciones hoy en día en todas las disciplinas de gran importancia para la sociedad como Ingeniería o Marketing y, en nuestro caso específico, Medicina, cuyas aplicaciones cabe destacar como antecedente histórico a partir de los años 70 (Rebasa, P. 2005).

Sus aplicaciones estudian el tiempo transcurrido entre dos eventos de interés. En Medicina, por ejemplo, la aparición de un acontecimiento adverso (en el presente proyecto recaída, metástasis o muerte) tras una intervención terapéutica o el tiempo entre el inicio de una infección y su diagnóstico (Armesto, Diana 2011). Debido a su antigüedad, existen gran cantidad de artículos científicos específicos que resaltan un interés cada vez mayor en este tema, con el objetivo de describir mediante modelos matemáticos una predicción fiable de tiempo de fallo o muerte de un “sujeto”, ya sea en sistemas biológicos o mecánicos.

Un alto porcentaje de estudios en el ámbito médico está especialmente relacionados con el cáncer, donde se intenta valorar la variable tiempo de ocurrencia de un evento (no

---

precisamente la muerte). A este tiempo se denomina “Supervivencia” (Rebasa,P. 2005). Cuando la ocurrencia del evento se presenta en todos los pacientes (no siendo esta situación una condición habitual), es necesario de técnicas estadísticas diferentes, debido a que las distribuciones que gobiernan las variables del tipo tiempo presentan unas características diferentes a las del resto de variables biológicas. Se aconseja entonces, planteamientos alternativos de análisis, ya que los convencionales se ven disminuidos tanto en precisión como en potencia, donde la media, desviación típica y la distribución normal no realizan un resumen exacto del tiempo de supervivencia (Gómez, Guadalupe 2004).

## **2.2 Teoría sobre el Análisis de Supervivencia**

Como hemos definido en los epígrafes anteriores, el termino supervivencia es una medida de tiempo a una respuesta, fallo, muerte o recaída de una determinada enfermedad o evento (Fernández, Pita 1995). Mencionamos también que este termino “supervivencia” se acuñó debido a que sus primeras aplicaciones estaban asociadas al evento de muerte de un paciente cuando se realizaban los análisis con este método.

En las enfermedades crónicas, tales como el cáncer, la supervivencia se evalúa a partir de la probabilidad de permanecer vivo durante un determinado periodo de tiempo. En el cáncer es de gran interés la medida de supervivencia en los primeros años, ya que son buenos indicadores tanto de la severidad de la enfermedad como el pronóstico; el cual se valora determinando el porcentaje de pacientes que sobrevive a menos de cinco años después del diagnóstico de la enfermedad (Fernández, Pita 1995). El interés del análisis se centra entonces en el tiempo que transcurre desde el diagnóstico hasta la muerte o recaída. En general, se dispondrá de un grupo de individuos al que se sigue un tiempo (t) hasta que se produzca el evento. Podemos entonces resumir en tres objetivos en los cuales se centra el análisis de supervivencia (Pérez, Santiago 1997):

1. Estimación: Calcular las probabilidades de observar un evento en determinado tiempo (t) a partir del origen.
2. Pruebas de hipótesis: Comparar experiencias entre dos o más grupos de observar el evento desde el origen.
3. Regresión: Estudiar y relacionar un conjunto de factores asociados a la supervivencia.

---

Para la realización de un análisis de supervivencia es necesario tener definido para cada individuo el instante de origen, el instante de ocurrencia del evento y por tanto el tiempo transcurrido hasta entonces (Kleinbaum, David G 2010). Sin embargo, antes de comenzar con cualquier análisis, es necesario comprender la necesidad de aplicar herramientas estadísticas diferentes a las convencionales las cuales son sesgadas e insuficientes para realizar una aproximación estadística verídica sobre el problema. El tiempo de supervivencia tiene una característica que lo hace inadecuado para otro tipo de análisis estadísticos y esto debido a la existencia de información truncada o individuos censurados (Martínez González, M. & de Irala, Estévez 2017). Por ello se debe introducir técnicas específicas para tratar esta información y poder controlar los sesgos.

### 2.2.1 Conceptos básicos

**Tiempo:** Es la variable de mayor importancia y representa la duración hasta que se produce un evento (muerte, abandono o recaída). Una definición formal puede darse de la siguiente manera: sea  $(T)$  una variable aleatoria positiva con una función de densidad  $f(t)$  y función de distribución de probabilidad  $F(t)$  donde (Kleinbaum, David G 2010):

$F(t) = P(T \leq t)$ ,  $0 < t < \infty$  Representa la probabilidad de ocurrencia del evento en un tiempo menor o igual que  $(t)$ .

**Función de supervivencia:** Se define como la probabilidad de que el evento de interés (muerte, recaída) ocurra más allá del tiempo  $(t)$ .

Notación matemática para  $S(t)$

$$S(t) = 1 - F(t) \Rightarrow F(t) = P\{T \leq t\}$$

**Función de riesgo (Hazard rate):** La función de riesgo  $h(t)$  representa la tasa instantánea de fallo; por tanto, la probabilidad condicionada por unidad de tiempo de que un evento ocurra en un instante  $t$ , dado que ha sobrevivido al tiempo  $t$  (Moore, D 2016).

Notación matemática para función de riesgo acumulada  $H(t)$

$$H(t) = P(T = t | T \geq t)$$

---

**Relación entre función de supervivencia y función de riesgo:** La función de supervivencia puede expresarse en función de la función de distribución acumulada  $S(t) = 1 - F(t)$ . A su vez, esta función puede expresarse con el cociente entre la función de densidad y la función de supervivencia  $h(t) = f(t) / S(t)$ .

### **El fenómeno de la censura**

El fenómeno de la censura ocurre cuando el sujeto no llega al evento de interés y se cuenta solo con un seguimiento parcial, no haciendo posible un acompañamiento completo dentro del estudio (Rivas, R, Pérez, M et al., 2014). A este suceso se le denomina “datos censurados”. En general existen tres razones para “Censurar” un dato:

1. Si un sujeto no presenta el evento antes del término del estudio.
2. Si no es posible seguimiento del sujeto durante el transcurso del estudio (pérdida de seguimiento).
3. Si una persona decide renunciar al estudio (o muere por distintas razones al evento de interés.)

Existen por ende, dos tipos de censura; la primera se denomina “*censura tipo I*” donde los individuos son observados hasta un tiempo determinado y la “*censura tipo II*” donde los individuos son observados hasta que ocurra un número determinado de eventos de interés.

Dentro de la “*censura tipo I*”, se presentan unos mecanismos necesarios de resaltar debido a que es el tipo de censura más frecuente en los estudios:

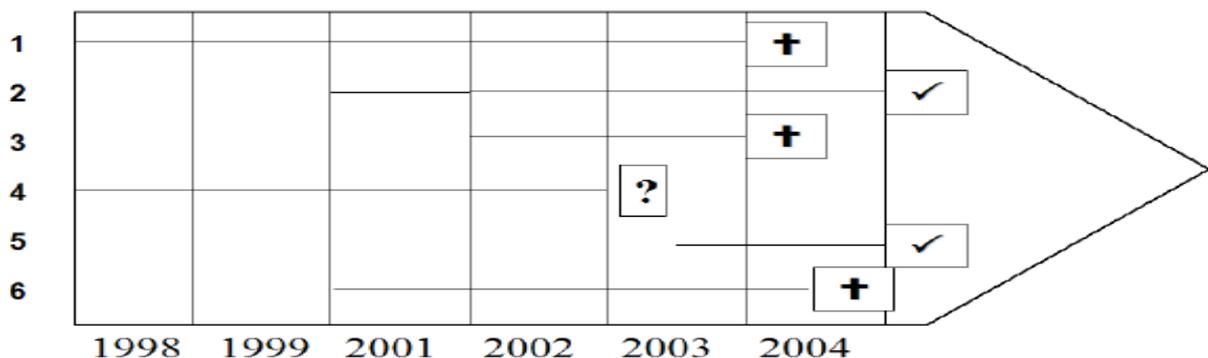
- Censura por la derecha: Se presenta cuando hasta la última observación que se realiza a un individuo, aún no ha ocurrido el evento que se desea observar y las relacionamos con las razones previamente comentadas.
- Censura por la izquierda: Es poco frecuente en un análisis de supervivencia, se presenta cuando en la primera observación que se realiza sobre un individuo, ya ha ocurrido el evento que se desea observar.

- Censura por intervalos: Se presenta cuando solo se sabe que al individuo le ocurre el evento de interés entre un instante  $T(i)$  y un instante  $T(j)$ .

Otro mecanismo que no hace posible la observación completa de los tiempos de seguimiento es el truncamiento. Existen dos mecanismos de truncamiento que pueden presentarse en los estudios:

- Truncamiento por la izquierda: Se presenta cuando el sujeto presenta una entrada tardía al estudio y comienza a observarse posteriormente al evento inicial.
- Truncamiento por la derecha: Se presenta cuando sólo se incluyen los individuos que presentan el evento de interés.

En la siguiente figura (Figura 2) se muestra un ejemplo de los tipos de censura en un estudio (López-Jiménez et al., 2005).



**Figura 2.** El esquema a modo de ejemplo, presenta 6 pacientes. Se han observado 3 muertes representados por cruces. Dos pacientes están censurados por seguir vivos al concluirse el estudio. El pacientes 4 también se encuentra censurado porque se perdió después de 4 años (López-Jiménez et al., 2005)

### **Métodos paramétricos**

El desarrollo del análisis de supervivencia utiliza dos métodos en los cuales se fundamenta para realizar los análisis estadísticos. Estos métodos son los denominados no-paramétricos y paramétricos.

De estos dos métodos, los no paramétricos suelen ser los más usados mediante el estimador de Kaplan-Meier y la comparación de curvas de supervivencia donde entraremos en más detalle más adelante (Kleinbaum, David G 2010).

Los métodos paramétricos consisten en aceptar que un determinado tipo de función de distribución paramétrica más conocida puede ser utilizada para describir nuestro modelo. Existen

---

varios tipos de funciones de distribución paramétrica: Distribución Gamma, del valor extremo, tipo Gumbel, Log-Normal y distribución Log-Logística. Sin embargo, en análisis de supervivencia las distribuciones más usadas con frecuencia son la distribución Exponencial y de Weibull. A continuación, profundizaremos en la definición matemática de estas dos distribuciones:

Distribución Exponencial: La función exponencial está definida por  $T \sim \exp(\lambda)$  con  $(\lambda) > 0$  donde  $f(t) = e^{-\lambda t}$ ,  $S(t) = e^{-\lambda t}$ ,  $h(t) = \lambda$  y  $H(t) = \lambda t$ . El modelo expresa la tasa de fallo en función de la combinación lineal de las  $n$  variables predictoras.

Distribución de Weibull: La función de Weibull está definida por  $T \sim W(\lambda, \alpha)$  con  $\lambda, \alpha > 0$  donde  $f(t) = \lambda \alpha (\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha}$ ,  $S(t) = e^{-(\lambda t)^\alpha}$ ,  $h(t) = \lambda \alpha (\lambda t)^{\alpha-1}$  y  $H(t) = (\lambda t)^\alpha - 1$ . Si  $\alpha = 1$  esta función es la exponencial, es decir la función exponencial es una particularización de la función más general de Weibull.

Estos modelos expresan la tasa de fallo instantáneo en función de las variables predictoras. Sin embargo, una de las condiciones que no permite a este método ser más utilizado es su rigidez, ya que los datos tienen que ajustarse a uno de los modelos. En el ámbito biomédico no suele ser común encontrarse datos que se ajusten a estos modelos, debido a que no se suele disponer de información suficiente para ajustar los datos a una distribución conocida, por lo que resulta de mayor utilidad la aplicación de métodos no paramétricos.

### **Métodos no paramétricos**

Como mencionábamos en el epígrafe anterior, estos métodos suelen ser lo más empleados, debido a que no son tan restrictivos cuando se realiza el análisis de supervivencia. Dentro de estas técnicas la más usada es el estimador de Kaplan-Meier (Moore, D 2016).

Estimador Kaplan-Meier: El objetivo principal de este estimador, es determinar la probabilidad de supervivencia en un grupo de sujetos en un intervalo de tiempo definido. La formulación matemática de la función de supervivencia en el tiempo ( $t$ ) viene determinada por:

$$\hat{S}(t) = \prod_{t_i \leq t} \left( \frac{n_i - d_i}{n_i} \right) \quad \forall t \geq t_1$$

donde

$d_i$  = número de eventos en el intervalo  $(t_{i-1}, t_i]$

$n_i$  = número de sujetos a riesgo al inicio del intervalo  $(t_{i-1}, t_i]$

---

Otro de los puntos fuertes del análisis de supervivencia y que supone un problema estadístico más frecuentes es la comparación de curvas de supervivencia. El problema viene cuando queremos comprobar si existe igualdad entre dos o más curvas de supervivencia estimadas a partir de *Kaplan-Meier* o por el contrario los grupos presentan diferencias estadísticamente significativas que pueden ser atribuidas al azar. Para esta comparación entre curvas de supervivencia, la técnica más usada es el *test de Log-Rank*; un estadístico Ji-Cuadrado donde se realiza una comparación entre datos observado (las curvas de supervivencia muestrales) y esperados (reuniendo todos los valores en una única muestra). Bajo la hipótesis nula ( $H_0$ ) se presenta el supuesto de la igualdad de dos o más funciones de supervivencia y para la hipótesis alternativa ( $H_1$ ), se presenta la que la existencia de diferencia estadísticamente significativa entre dos o más grupos (Borges, 2017)

### **Modelo de semiparamétrico de Cox**

Si deseamos evaluar el efecto de múltiples variables sobre la función de supervivencia, es necesario considerar otros métodos que permitan crear un modelo que describa la contribución de estas variables explicativas en la probabilidad de supervivencia. El modelo de Cox(1972), es un modelo semiparamétrico, debido a que incluye en su análisis una parte paramétrica y otra no paramétrica. Una definición matemática define el riesgo para i-ésimo individuo mediante la siguiente expresión:

$\lambda(t; Z_i(t)) = \lambda_0(t) e^{\beta Z_i(t)}$  Donde  $Z_i(t)$  es el vector de covariables para el i-ésimo individuo en el tiempo ( $t$ ). La parte paramétrica  $ri(t) = e^{\beta Z_i(t)}$ , llamada puntaje de riesgo (risk score), donde  $\beta$  es el vector de parámetros de la regresión, la cual esta claramente especificada y describe los efectos relativos de los parámetros de regresión estimados sobre el riesgo. La parte no paramétrica es el  $\lambda_0(t)$ , la función de riesgo base, que es una función arbitraria y no especificada (Borges, 2017).

El modelo de regresión de Cox se conoce como el modelo de riesgos proporcionales, contiene el riesgo para dos sujetos con el mismo vector de covariables, siendo constante en el tiempo, es decir:

$$\frac{\lambda(t; Z_i(t))}{\lambda(t; Z_j(t))} = \frac{\lambda_0(t) e^{\beta Z_i(t)}}{\lambda_0(t) e^{\beta Z_j(t)}} = \frac{e^{\beta Z_i(t)}}{e^{\beta Z_j(t)}}$$

---

Suponiendo que la muerte ha ocurrido en el tiempo  $(t)^*$ , entonces la verosimilitud de que la muerte le ha ocurrido al individuo  $i$ -ésimo y no al otro individuo es:

$$L_i(\beta) = \frac{\lambda_0(t^*)r_i(t^*)}{\sum_j Y_j(t^*)\lambda_0(t^*)r_j(t^*)} = \frac{r_i(t^*)}{\sum_j Y_j(t^*)r_j(t^*)}$$

El producto de los términos de la expresión anterior  $L\beta = \prod L_i(\beta)$  es llamada la verosimilitud parcial. La maximización del  $\text{Log}(L(\beta))$  da una estimación para  $\beta$  sin necesidad de estimar el parámetro de ruido o función de riesgo base  $\lambda_0(t)$ .

### 2.3 Planteamiento del problema

En este epígrafe, se presenta un esbozo del planteamiento a desarrollar en los epígrafes a continuación. Este Capítulo 2 desarrolla en su primera parte una introducción teórica al análisis de supervivencia y sus antecedentes relacionados con los análisis estadísticos utilizados. Consiguientemente desarrolla una presentación de las variables dentro del estudio y una preparación de las mismas para que cumplan con los aspectos básicos de análisis de supervivencia y los requerimientos mínimos para su aplicación mediante el programa R. Una vez preparados los datos, el primer análisis contempla un desarrollo descriptivo para las variables categóricas. En cualquier estudio sobre datos reales, un análisis estadístico debe comenzar con una descripción univariada y completa de los datos. Es por eso que se realiza una descripción univariante de los datos observados mediante la estimación de la función de supervivencia para las situaciones de recaída y muerte contemplando la variable género. Consiguientemente se analizan las diferencias de curvas de supervivencia entre los grupos dentro de las variables categóricas. Se evalúa simultáneamente el efecto de las variables explicativas sobre las situaciones de recaída y muerte. Se ajusta el mejor modelo mediante el método semiparamétrico de Cox y por último se evalúa el modelo mediante un método paramétrico. Al final, se presenta en este Capítulo 2 un análisis de resultados obtenidos globalmente mediante los diferentes análisis.

---

### 2.3.1 Descripción de la base de datos

Se trata de una serie de 99 pacientes intervenidos quirúrgicamente y diagnosticados de carcinoma de células renales (CCR). El seguimiento del estudio comprende el periodo de tiempo aproximado entre los años 2001 - 2014. Las variables consideradas en el trabajo se detallan en la siguiente tabla (Tabla 1):

**Tabla 1: Variables de estudio**

Variables	Descripción
Id	Identificador único para cada paciente
Sexo	Género (Hombre/Mujer)
Edad	Edad en la fecha de cirugía
Tipo de tumor	Tipo de carcinoma que padece el paciente
Estadio	Estadificación del cáncer según el sistema TNM
Grado histológico (Führman)	Sistema de Clasificación del carcinoma mediante pruebas nucleares.
Fecha de cirugía	Fecha de la primera intervención
Primera recidiva o primer evento de progresión(Local/Metástasis):	Tipo de evento de recaída
Fecha	Fecha del primer evento de recaída
Intervalo libre de enfermedad	Rango en días del paciente libre de la enfermedad
Exitus por la enfermedad	Exitus letalis o muerte por la enfermedad
Exitus por otras causas	Exitus letalis o muerte por causas diferentes a la enfermedad
Supervivencia global	Ultima fecha de seguimiento
Quimioterapia	Inicio: fecha de inicio del tratamiento Fin : Fin del tratamiento Tratamiento: Tipo de tratamiento
Radioterapia	Inicio: Fecha de inicio de tratamiento Fin: Fin de tratamiento Dosis: Dosis aplicada de tratamiento

---

### 2.3.2 Codificación de variables

Para el desarrollo del análisis estadístico en R, es necesaria la codificación y creación de algunas de las variables dentro de la base de datos.

**Sexo** : 1= Hombre ; 0 = Mujer (**Codificación binaria**)

**Tipo de tumor** : Simplificación de variables según tipo de cáncer renal (Cancer de riñón en España, 2017) (**Codificación en grupos**)

CCC : Carcinoma Renal de Células Claras

CRP : Carcinoma Renal Papilar

M + CRP : Metastasis Carcinoma Renal Papilar

M + CCC : Metastasis Carcinoma Células Claras

M + A + CCC : Metastasis Adenocarcinoma Células Claras

M + A + CRP : Metastasis Adenocarcinoma Células Renal Papilar

CCC + CRS : Carcinoma Células Claras con áreas Sarcomatoides

CRC + CRS : Carcinoma de Células Renales Cromóforo con áreas Sarcomatoides

CRJ : Carcinoma de células renales inclasificable

CCR: Carcinoma cromóforo

**Tiempo\_Rec** : Años[Fecha de la cirugía - Fecha primer evento recaída] (**Nueva variable**)

**Tiempo\_Sup** : Años [Fecha de la cirugía - Supervivencia global] (**Nueva variable**)

**Evento\_1** : 1= Recidiva ; 0 = Censura (**Nueva variable**)

**Evento\_2** : 1 = Exitus por enfermedad ; 0 = Censura (**Nueva variable**)

**Quimioterapia** : 1 = Si ; 2 = No (**Codificación binaria**)

**Radioterapia** : 1 = Si ; 2 = No (**Codificación binaria**)

**Edad**: Rango de edades: [< 50], [50 - 60], [60-70],[>70 años]\* (**Codificación en rangos**)

\* La incidencia máxima de CCR se presenta entre 60 - 70 años y predominan los casos de hombres sobre las mujeres. Algunos casos en pacientes <40 años es objeto de estudio en la actualidad y se relaciona con antecedentes familiares. Se ajusta un rango de edad según vulnerabilidad de incidencia de CCR. Para transformar esta variable en nuestro estudio nos apoyamos en la bibliografía (de Vicente, JM Gómez et al., 2003) (Ljungberg, B, Cowan, N et al., 2010) (Guillermo, Gállego-Sales, Jamaica-Verduzco, E et al., 2010) (Gomez Perez, L, Budia, A et al., 2007)

## 2.4 Análisis estadístico

### 2.4.1 Resumen estadístico para variables continuas y factores

<b>Sexo</b>					
Hombre		63			
Mujer		36			

<b>Edad en fecha de cirugía</b>		<b>Estadio</b>	
Min	29	pT1b pNx	9
1st Qu.	52	pT1aNx Mx	4
Median	61	pT1apNx pMx	4
Mean	60.77	pT1bNx Mx	4
3rd Qu.	69.50	(Other)	63
Max.	87	NA's	15

<b>Intervalo días sin enfermedad</b>		<b>Tiempo_Rec</b>		<b>Tiempo_Sup</b>	
Min.	38	Min.	38	Min.	13
1st Qu	492.8	1st Qu	492.8	1st Qu	960
Median	1525.01	Median	1550	Median	1789
Mean	1723.4	Mean	1725.9	Mean	1990
3rd Qu.	2956.2	3rd Qu.	2960.5	3rd Qu.	2969
Max.	4663	Max.	4663	Max.	5055
NA's	7	NA's	8	NA's	2

Tabla 2: Resumen descriptivo de variables de estudio

En la tabla 2 se presenta un resumen descriptivo de las variables continuas y factores dentro de los datos de pacientes con CCR. De esta tabla se pueden extraer dos datos poblacionales de interés: El primer dato, esta relacionado con la variable sexo, podemos apreciar que en el estudio los hombres presentan mayor afección que las mujeres para CCR (Ljungberg, B, Cowan, N et al., 2010). Como segundo dato relevante, nos enfocamos en la variable edad, donde nuestro media de edad es aproximadamente 61 años y guarda mucha relación con los valores estudiados en la bibliografía, donde sitúa la población más vulnerable a esta enfermedad entre 60 - 70 años (Gomez Perez, L, Budia, A et al., 2007).

### Tipo de tumor relacionado con Sexo

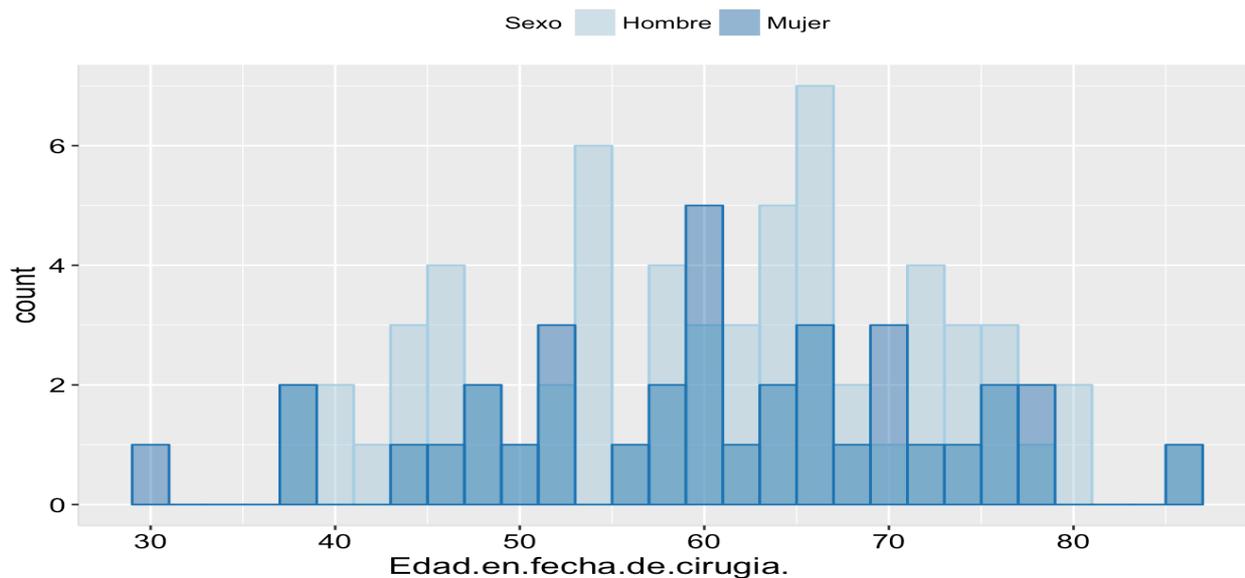
**Tabla 3 : Relación de variables tipo de tumor y sexo**

<i>Tipo.de.tumor</i>	<i>Sexo</i>		<i>Total</i>
	Hombre	Mujer	
CCC	39 60.9 %	25 39.1 %	64 100 %
CCC + CRS	4 80 %	1 20 %	5 100 %
CCR	0 0 %	2 100 %	2 100 %
CRC + CRS	0 0 %	1 100 %	1 100 %
CRI	1 100 %	0 0 %	1 100 %
CRP	12 66.7 %	6 33.3 %	18 100 %
M + A + CCC	2 100 %	0 0 %	2 100 %
M + A + CRP	1 100 %	0 0 %	1 100 %
M + CCC	3 75 %	1 25 %	4 100 %
M + CRP	1 100 %	0 0 %	1 100 %
<i>Total</i>	63 63.6 %	36 36.4 %	99 100 %

$\chi^2=9.182 \cdot df=9 \cdot Cramer's V=0.305 \cdot Fisher's p=0.562$

La tabla 3 analiza la distribución de los diferentes tipos de tumores en relación con la variable sexo. Es importante destacar el objetivo de esta comparación. En un artículo reciente publicado en la revista Cell Metabolism el cual se titula “Sex and Gender: Critical Variables in Pre-Clinical and Clinical Medical Research” se resalta la importancia de incorporar la variable sexo en la investigación preclínica y clínica con el objetivo de mejorar el sesgo y la comprensión de los mecanismos por los cuales los procesos metabólicos difieren de esta condición. El artículo explica que «existe una tendencia en las investigaciones a tratar a ambos como equivalentes sin tener en cuenta como afectan las fluctuaciones de los diferentes niveles de hormonas sexuales en los resultados experimentales, lo que limita la comprensión de los mecanismos moleculares que difieren los dimorfismos sexuales» (Morselli, E, Aaron, F et al., 2016). Es por este motivo que se destacará esta variable “sexo” en los análisis de los siguientes epígrafes.

En la tabla 3 se destacan los tipos de tumores *Carcinoma de Células Claras (CCC)* y el *Carcinoma Renal Papilar (CRP)* como los más recurrentes. Para el CCC con 39 casos, donde el 60% de los casos se presenta en hombres y 39% en mujeres. El CRP con 18 casos diagnosticados donde un 66% de los casos lo representan en hombres y 33% las mujeres. De manera global de un total de 99 pacientes, los hombres ocupan un 63% frente a un 36% de mujeres en casos de diagnóstico de CCR. Vemos entonces, como de forma absoluta y relativa, los diagnósticos sitúan a los hombres a diferencia de las mujeres como la población más vulnerable a padecer este tipo de cáncer.



**Figura 3 : Distribución edad**

La Figura 3 resume en un histograma la distribución de la variable edad en relación el sexo dentro de la base de datos. Vemos como los casos de diagnóstico en este tipo de cáncer aumenta entre los 50 - 70 años. Este dato contrasta perfectamente con la población más vulnerable en la bibliografía (Guillermo, Gállego-Sales, Jamaica-Verduzco, E et al., 2010). Aún así, cabe resaltar que esta enfermedad es objeto de estudio en pacientes menores de 40 años debido a su rareza, las cuales atribuyen a factores genéticos y hereditarios la aparición de esta afección en pacientes jóvenes (Gomez Perez, L, Budia, A et al., 2007). Es por eso que notamos una ausencia en la frecuencia del histograma para este rango de edad (30-40 años).

### **Distribución de edad según sexo en variable edad recodificada**

**Tabla 4 : Distribución de la edad según sexo (variable recodificada)**

<i>Edad.en.fecha.de.cirugía.</i>	<i>Sexo</i>		<i>Total</i>
	Hombre	Mujer	
<50 años	28 66.7 %	14 33.3 %	42 100 %
50-60 años	16 61.5 %	10 38.5 %	26 100 %
60-70 años	18 62.1 %	11 37.9 %	29 100 %
>70 años	1 50 %	1 50 %	2 100 %
<i>Total</i>	63 63.6 %	36 36.4 %	99 100 %

$\chi^2=0.408 \cdot df=3 \cdot Cramer's V=0.064 \cdot Fisher's p=0.956$

En la tabla 4 se realizó una recodificación de la variable por intervalos edad. El objetivo de esta recodificación es poder analizar posteriormente la supervivencia por grupo de edad y sexo. Consiguientemente, centrarnos en el rango de edad de mayor vulnerabilidad según la bibliografía. Podemos observar que en este rango entre 50-70 años, se centran el 56% de casos de diagnóstico de CCR.

---

## 2.4.2 Estimación de función de supervivencia para variables de tiempo

Se analiza la primeras variables de respuesta y censura. Esta primera situación presenta el tiempo desde la operación hasta la muerte del paciente.

### Situación 1:

Tiempo = Tiempo de supervivencia  
evento = Exitus por enfermedad  
Grupo = Género

Iniciamos con un resumen descriptivo de supervivencia para la variable de censura “**Envento\_2**”.

0 = 54  
1 = 43  
Na= 2

[ 1]	14	13+	13+	2	7+	11+	7	7	4	8+	10+	10+	10+	10+	8	9+	10+
[18]	9+	9+	9+	9+	9+	9	9+	5	9+	9+	4+	8+	9+	6	8+	7	8+
[35]	8+	4	8+	8+	7+	4+	4+	3+	7	8+	8+	7	7+	2+	8+	8	0+
[52]	8+	8	5	2+	2+	1	1	6+	7+	1	3+	4+	4	5	6+	6+	6+
[69]	4	5	5	2	3	2	5	4+	3	3	4	2	1	1+	1	3+	1
[86]	1	1+	0+	0+	3	1	3	3	3	3	1	NA	2+	NA			

De los 99 individuos, 43 han sufrido exitus por la enfermedad, 54 son observaciones censuradas por la derecha (+) y se presentan 2 datos perdidos (NA).

### Función de supervivencia mediante el estimador de Kaplan-Meier para la situación

#### 1

Se analiza la función de supervivencia mediante el método no paramétrico de Kaplan-Meier. Las curvas de Kaplan-Meier son útiles para visualizar la diferencia de la supervivencia entre grupos categóricos. Sin embargo, se reduce su funcionalidad cuando intentamos evaluar el efecto de variables cuantitativas tales como la edad, expresión genética, etc (Rao, PV et al., 2000).

Para una mejor interpretación en el análisis, se realizó una conversión de días a años sobre las variables de tiempo. También se transformó la variable edad en rangos para visualizar la curva de supervivencia según el grupo de edad (A Torrelles, M Garc et al., 2006).

---

## Supervivencia global situación 1

```
2 observations deleted due to missingness
  n  events  median 0.95LCL 0.95UCL
97   43     8      7      NA
```

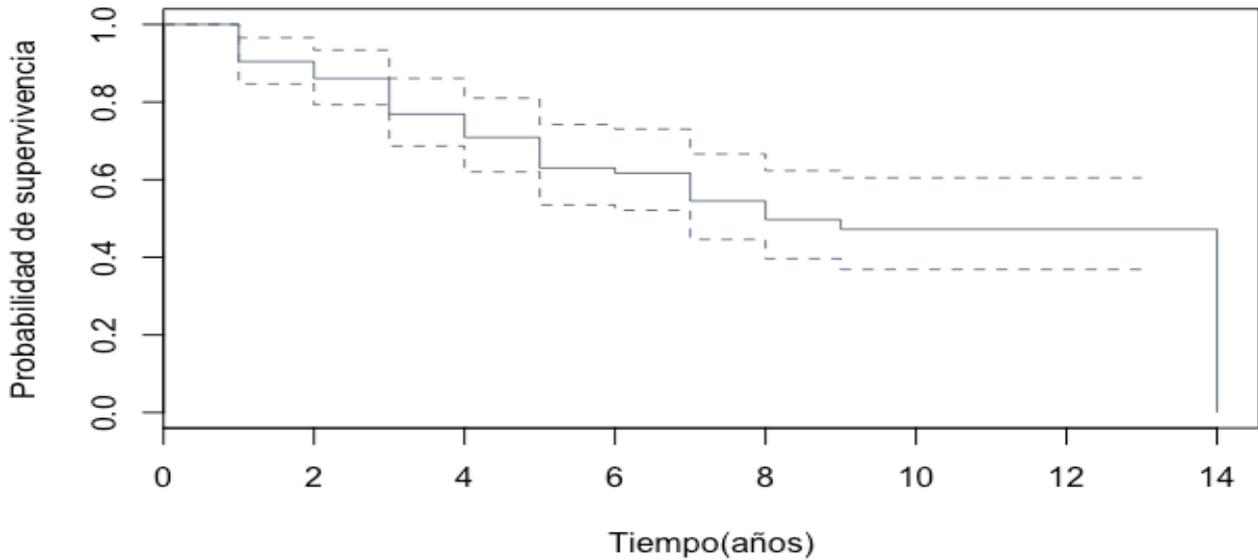
La función `survfit()` en R, nos muestra un breve resumen descriptivo de supervivencia. Obtenemos que el número total de pacientes evaluados es de 97, el número de eventos que presentaron exitus por enfermedad es de 43 y la media de supervivencia es de 8 años.

Podemos obtener resumen más detallados del objeto de supervivencia. Este resumen nos muestra el tiempo de ocurre algún evento, el número de individuos que estaba en riesgo en ese momento, el número de eventos y la estimación de supervivencia con su intervalo de confianza al 95%.

```
2 observations deleted due to missingness
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1    94     9   0.904  0.0303    0.847    0.966
  2    83     4   0.861  0.0359    0.793    0.934
  3    75     8   0.769  0.0444    0.687    0.861
  4    64     5   0.709  0.0483    0.620    0.810
  5    54     6   0.630  0.0526    0.535    0.742
  6    48     1   0.617  0.0531    0.521    0.730
  7    43     5   0.545  0.0558    0.446    0.666
  8    34     3   0.497  0.0574    0.396    0.623
  9    20     1   0.472  0.0596    0.369    0.605
 14     1     1   0.000    NaN      NA      NA
```

A partir del resumen anterior, obtenemos el gráfico de la función global de supervivencia para esta **Situación 1**.

## Función de supervivencia Kaplan-Meier



**Figura 4 : Función global de supervivencia**

En la Figura 4 se puede ver cómo evoluciona la probabilidad de supervivencia a lo largo del tiempo en esta primera situación. Podemos destacar entre los 5 y 7 años una situación estacionaria, después la supervivencia comienza a descender de forma más lenta. Al contrastar los valores del gráfico con el resumen, notamos que a los 5 años la tasa de supervivencia 63%, en media, el 60 % de los pacientes de este estudio siguen vivos, el número de individuos en riesgo es de 54 han ocurrido 6 eventos.

### Supervivencia por género

En este apartado se analiza la función de supervivencia en relación con la variable explicativa género.

```

2 observations deleted due to missingness
              n events median 0.95LCL 0.95UCL
data_CCR$Sexo=Hombre 61    31     7      5     NA
data_CCR$Sexo=Mujer  36    12    14     9     NA
    
```

Al observar la probabilidad de supervivencia separada por grupos dentro de la variable género, notamos que el número de exitus por enfermedad es mayor en el grupo de hombres donde se presentan 31 eventos a diferencia de los 12 eventos presentados por el grupo de las mujeres. A pesar de que el estudio presenta una proporción mayor de hombres, la bibliografía aplicada a casos similares (StataCorp, LP 1985), prevé que efectivamente existe una mayor influencia de casos de muerte para este grupo. Otro dato importante se encuentra relacionado

con la supervivencia, donde las mujeres con una media de 14 años, presentan una mejor supervivencia que los hombres con una media de 7 años ante esta enfermedad.

2 observations deleted due to missingness

data_CCR\$Sexo=Hombre							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper 95% CI
1	59	8	0.864	0.0446		0.781	0.956
2	50	3	0.813	0.0510		0.719	0.919
3	44	4	0.739	0.0582		0.633	0.862
4	37	4	0.659	0.0642		0.544	0.797
5	30	4	0.571	0.0690		0.451	0.724
6	26	1	0.549	0.0698		0.428	0.704
7	25	4	0.461	0.0711		0.341	0.624
8	19	3	0.388	0.0712		0.271	0.556

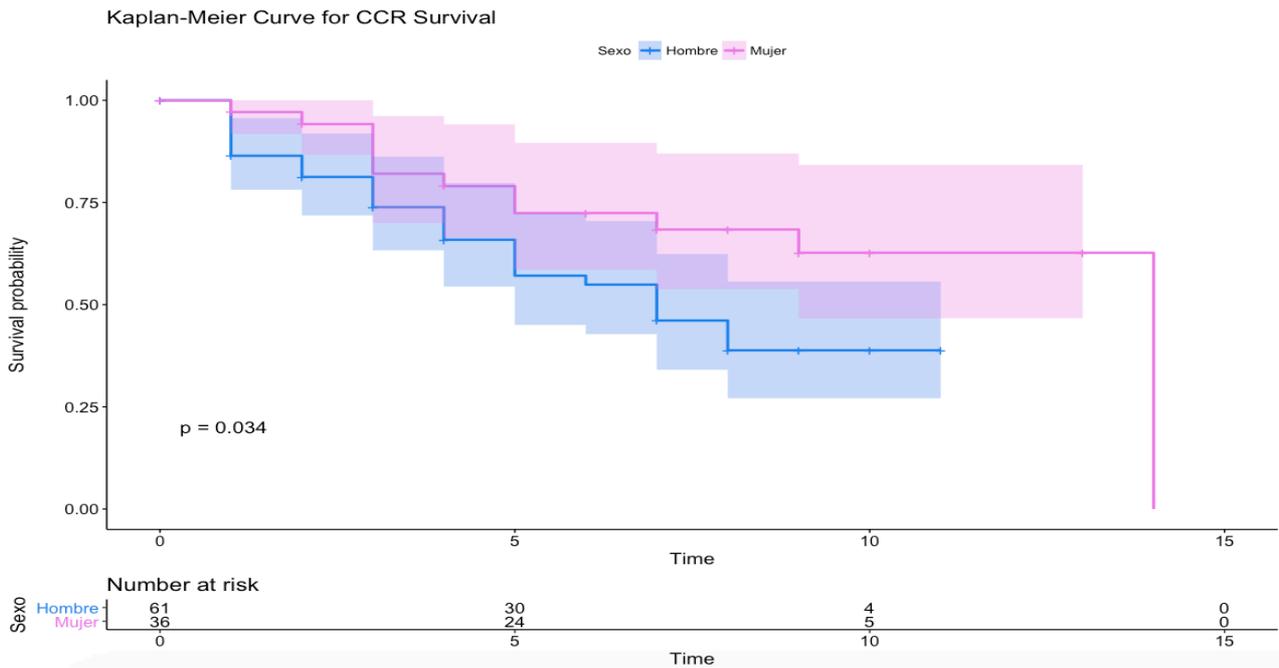
  

data_CCR\$Sexo=Mujer							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper 95% CI
1	35	1	0.971	0.0282		0.918	1.000
2	33	1	0.942	0.0398		0.867	1.000
3	31	4	0.820	0.0665		0.700	0.962
4	27	1	0.790	0.0706		0.663	0.941
5	24	2	0.724	0.0786		0.585	0.896
7	18	1	0.684	0.0839		0.538	0.870
9	12	1	0.627	0.0943		0.467	0.842
14	1	1	0.000	NaN		NA	NA

Al separar la función de supervivencia por grupos, los resúmenes descriptivos nos aportan la información para cada grupo dentro de la variable. El objetivo de esta división es comparar curvas de supervivencia entre el grupo de hombres y mujeres.

podemos observar que efectivamente el numero de exitus por CCR es menor en las mujeres, donde la supervivencia en los primeros 5 años es la mitad que para los hombres.

Consiguientemente se genera el gráfico para la comparación de curvas de supervivencia entre estos dos grupos.



**Figura 5 : Función de supervivencia por género**

Cuando las curvas visualmente no marcan diferencias, podemos recurrir al test de *Long - Rank*, donde planteamos las hipótesis de diferencia entre grupos y decidimos con un nivel de significación de 95%, si existe diferencia estadísticamente significativa entre curvas de supervivencia.

En la Figura 5 las curvas parecen ir muy paralelas durante los primeros 5 años, luego la curva para el grupo de los hombres presenta un punto de inflexión y desciende más rápidamente. Esta comparación la evaluaremos a partir de un test *Long Rank* para observar si efectivamente las curvas presentan diferencias significativas.

Planteamos las hipótesis para el contraste de comparación entre curvas:

$$H_0: S_H(t) = S_M(t)$$

$$H_1: S_H(t) \neq S_M(t)$$

Mediante la función `ggsurvplot()` de R, indicamos que nos muestre en la Figura 5 el p-valor para la comparación entre grupos. Con un *P-valor* = 0.034 ajustado en la gráfica, rechazamos la hipótesis de igualdad y por tanto podemos decir con un nivel de significación del 95% que existe una diferencia estadísticamente significativa.

Ahora analizamos la **Situación 2**. Las variables respuesta y censura están relacionadas con el periodo de tiempo entre la operación y la recidiva o recaída del paciente en la enfermedad.

Situación 2

Tiempo = Tiempo hasta recaída  
 evento = recaída  
 Grupo = Género

Iniciamos con un resumen descriptivo de supervivencia para la variable de censura **“Envento\_1”**.

0 = 71  
 1 = 21  
 Na= 7

```
[1] 9 13+ 13+ 1 7+ 11+ 4 5+ NA 8 10+ 10+ 10+ 10+ 4+ 9+ 10+
[18] 9+ 9+ 9+ 9+ 9+ 7+ 9+ 4+ 9+ 9+ 4+ 8+ 9+ 1 8+ 2+ 8+
[35] 8+ 3+ 8+ 8+ 7+ 4+ 4+ 3+ 5 8+ 8+ 5 7+ 2+ 8+ 2+ 0
[52] 8+ 4+ 3+ 2+ 2+ 0 0 6+ 7+ 1 3+ 4+ 2+ 1+ 6+ 6+ 6+
[69] 0 0+ 3 1+ 1 2 1+ 4+ 1 3 2+ 1 0 NA 1 1+ 1+
[86] 1 NA+ NA NA 2+ 1+ 0+ 1+ 2+ 0+ 0+ NA+ NA+ NA+
```

De los 99 individuos, 21 individuos han sufrido recaída, 71 son observaciones censuradas por la derecha (+) y 7 datos perdidos(NA).

```
8 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
91 21 NA NA NA
```

Aplicamos la función survfit() sobre la situación de recaída y no obtenemos todos los valores para el descriptivo de supervivencia. Esta situación se presenta porque la curva de supervivencia no alcanza la línea del 50% antes del final del estudio, es por eso que la estimación de la mediana de supervivencia es infinita y el resultado es NA (Dalgaard, P 2008).

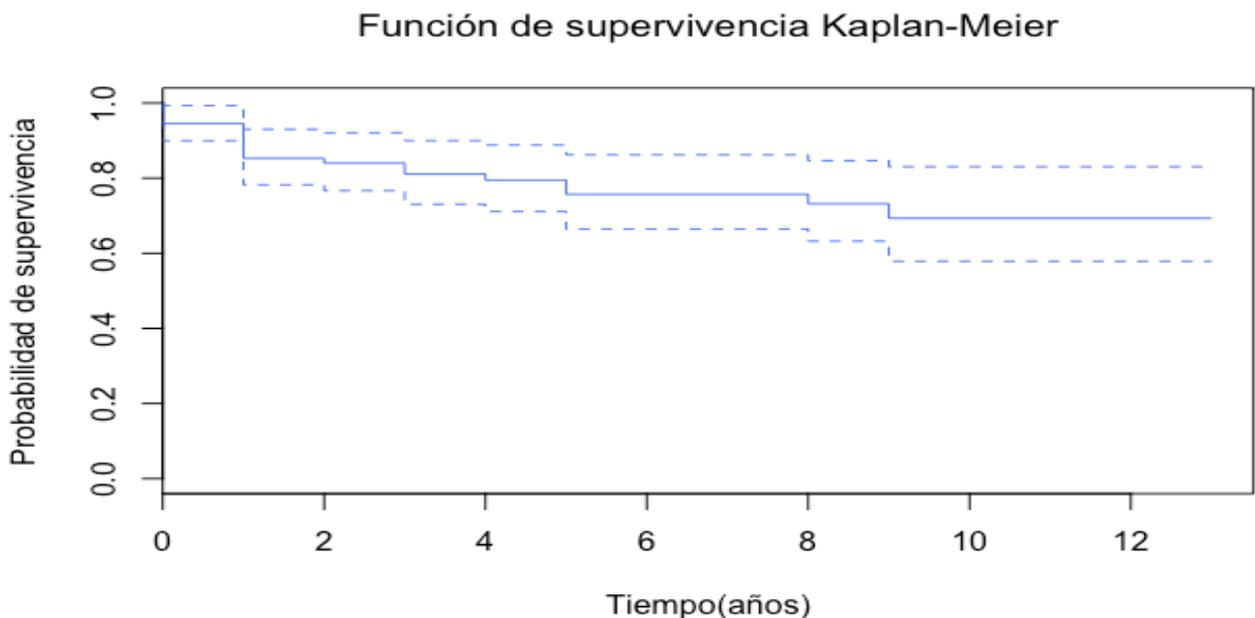
**Supervivencia global situación 2**

```
8 observations deleted due to missingness
time n.risk n.event survival std.err lower 95% CI upper 95% CI
0 91 5 0.945 0.0239 0.899 0.993
1 82 8 0.853 0.0377 0.782 0.930
2 67 1 0.840 0.0393 0.767 0.921
3 57 2 0.811 0.0431 0.730 0.900
4 51 1 0.795 0.0451 0.711 0.888
```

5	42	2	0.757	0.0502	0.665	0.862
8	30	1	0.732	0.0545	0.632	0.847
9	19	1	0.693	0.0638	0.579	0.830

El resumen descriptivo más detallado para la *situación 2* sobre la supervivencia global marca una probabilidad de recaída muy leve. A los 9 años la probabilidad de recaída no alcanza aún el 50%. Pero es necesario resaltar que este caso presenta un 77% de datos censurados, cantidad considerable. esto limita y sesga un la función estimada de supervivencia (Dalgaard, P 2008).

Ahora a partir de los datos anteriores, se realiza el gráfico de función de supervivencia global para esta *situación 2*.



**Figura 6 : Función de supervivencia global para evento recaída**

En la Figura 6 observamos que la curva de supervivencia para el evento recaída sufre solo un descenso escalonado en los primero dos años, después, se mantiene estable y su tendencia al infinito puede ser explicada por la gran cantidad de datos censurados (Dalgaard, P 2008).

---

## Supervivencia por género

Estimamos la función de supervivencia para la *situación 2*, y la relacionamos con la variable explicativa género. Consiguientemente evaluamos mediante un test *Long Rank* si existe una diferencia significativa para los grupos dentro de la variable género.

```
8 observations deleted due to missingness
              n events median 0.95LCL 0.95UCL
data_CCR$Sexo=Hombre 58      16      NA      NA      NA
data_CCR$Sexo=Mujer  33       5      NA      NA      NA
```

Para este breve resumen descriptivo podemos resaltar que a pesar de no tener los valores para la mediana e intervalos de confianza, la situación de recaída se encuentra en mayor proporción para el grupo de hombres con 16 eventos en comparación con los 5 eventos presentados por el grupo de las mujeres. De manera más simple, el porcentaje de recaída es de 27% para el total de los hombres y del 15% para las mujeres.

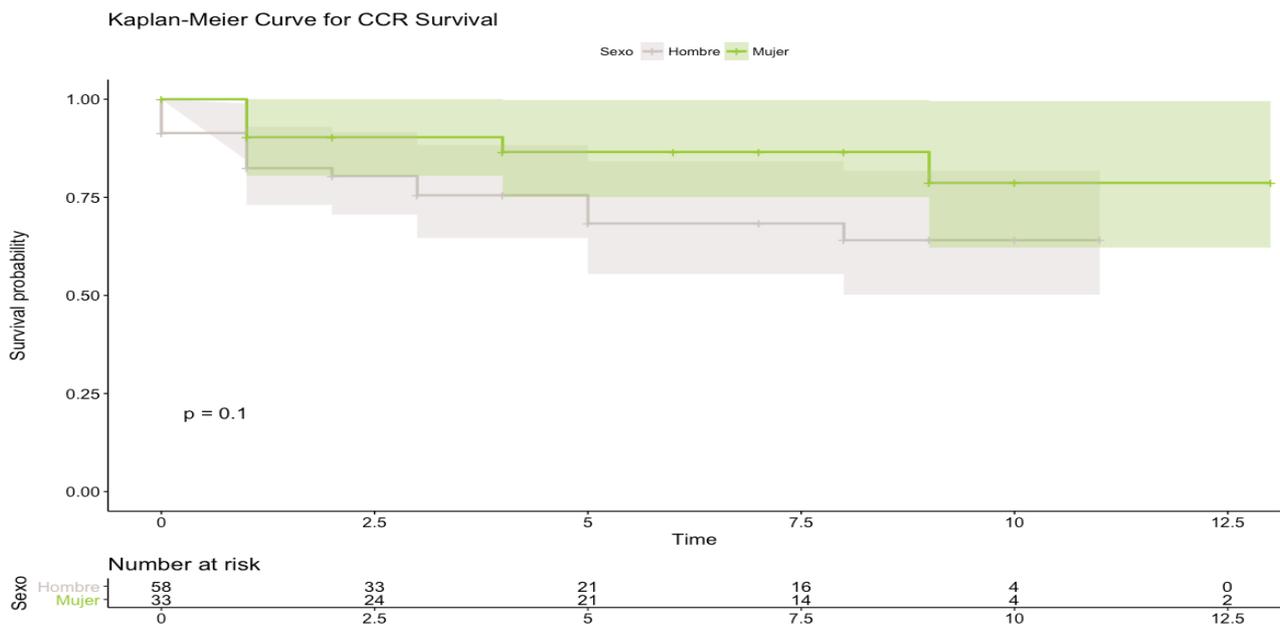
Obtenemos un resumen descriptivo más detallado para este situación 2.

```
8 observations deleted due to missingness
              data_CCR$Sexo=Hombre
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    58     5    0.914  0.0369    0.844    0.989
  1    51     5    0.824  0.0505    0.731    0.929
  2    41     1    0.804  0.0531    0.706    0.915
  3    33     2    0.755  0.0601    0.646    0.883
  5    21     2    0.683  0.0728    0.555    0.842
  8    16     1    0.641  0.0798    0.502    0.818

              data_CCR$Sexo=Mujer
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1    31     3    0.903  0.0531    0.805    1.000
  4    24     1    0.866  0.0628    0.751    0.998
  9    11     1    0.787  0.0943    0.622    0.995
```

En el resumen observamos como la probabilidad de recaída se mantiene estable en el grupo de las mujeres a diferencia del grupo de los hombres.

Evaluamos mediante el gráfico de supervivencia si visualmente podemos apreciar una diferencia en las curvas entre grupos.



**Figura 7 : Función de supervivencia por género para evento de recaída**

La Figura 7 representa la curva de supervivencia para la *situación 2*. Las curvas siguen una tendencia muy similar. Para la valoración de diferencias entre las curvas utilizaremos el test de *Long Rank*.

Planteamos las hipótesis para el contraste de comparación entre curvas:

$$H_0: S_H(t) = S_M(t)$$

$$H_1: S_H(t) \neq S_M(t)$$

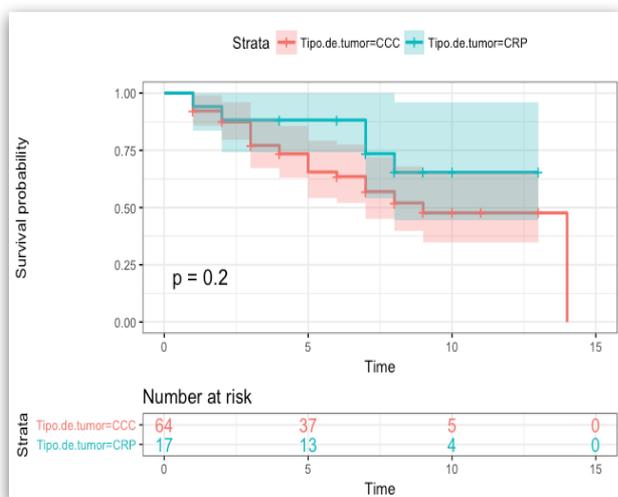
Con un *P-valor* = 0.1 ajustado en la gráfica, rechazamos la hipótesis de diferencia y por tanto podemos decir con un nivel de significación del 95% que no existe una diferencia estadísticamente significativa entre curvas de supervivencia. Observamos que la curva no llega al 50% de probabilidad acumulada para ninguno de los grupos, esto indica que menos del 50% de casos sufren recaída.

### 2.4.3 Comparación de curvas de supervivencia en variables explicativas

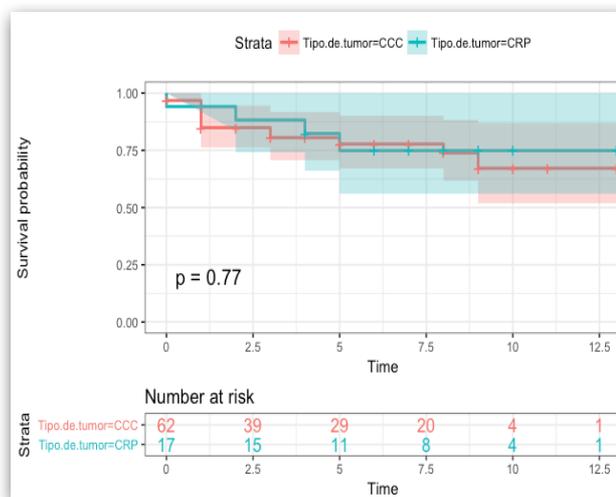
Además de la variable “Sexo”, el estudio presenta otro tipo de variables explicativas cuya influencia sobre los pacientes puede ser considerada dentro del estudio. En este apartado se propone realizar un análisis univariante mediante el método no paramétrico de *Kaplan-Meier* para estas variables.

La finalidad de este apartado es comprender cómo se ve afectada función de supervivencia en los diferentes grupos dentro de las variables categóricas. Es necesario especificar que usamos como guía ejemplos teóricos de análisis similares (Moore, D 2016). En la variable categórica (*tipo de tumor*) evaluamos la influencia de los tipos de tumores más recurrentes en el estudio y a partir de su función de supervivencia analizamos si existe una diferencia significativa entre los grupos. Por otra parte, usaremos la variable categórica codificada de rango de edad con tal de evaluar el rango que presenta menos probabilidad de supervivencia. Por ultimo, evaluamos los efectos que tiene haber recibido quimioterapia y (o) radioterapia y como se ve afectada la función de supervivencia con estas variables.

### Función de Supervivencia por tipo de tumor y sexo.



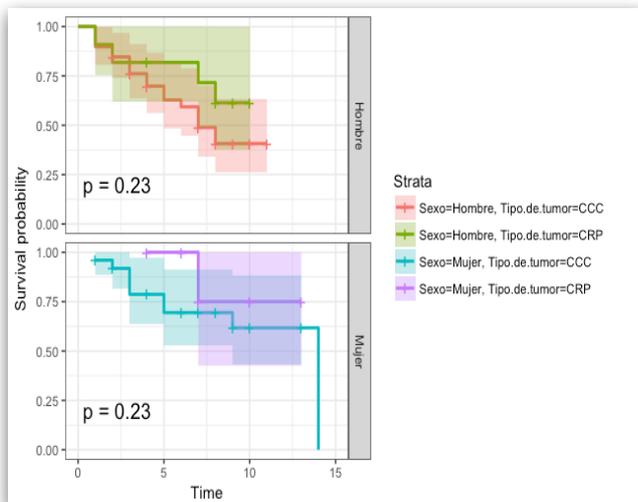
**Figura 8:** Función de supervivencia t. de tumor *Situación 1* (Muerte)



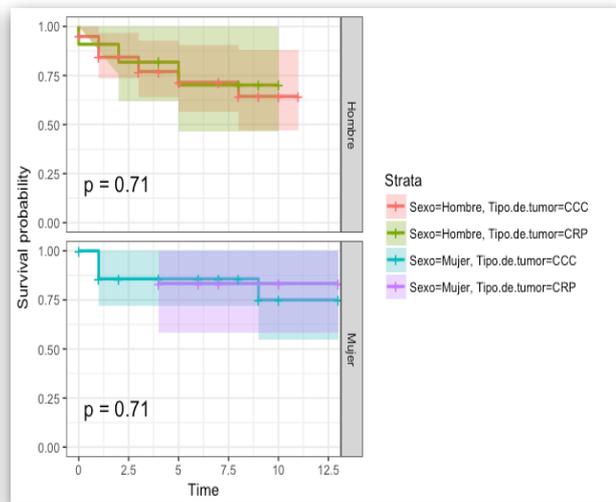
**Figura 9:** Función de supervivencia tipo de tumor *Situación 2* (Recaída)

En este apartado se realiza un análisis sobre las dos situaciones en conjunto. Se analiza la influencia del los tipos de tumores más recurrente en el estudio (CCC,CRP) sobre los eventos de exitus por enfermedad y recaída. Podemos observar en la Figura 8 como la supervivencia en los primeros años es muy similar, sin embargo los pacientes con CRP tienen mejor pronóstico en cuanto a probabilidad de supervivencia una vez superado los dos años. Para la Figura 9 observamos que la probabilidad de recaída es igual para los dos tipos de cáncer. Estas conclusiones se pueden percibir visualmente, pero aplicando el test de Long-Rank donde se tienen en cuenta todos los tiempos para analizar la supervivencia de ambos grupos nos dirá si efectivamente existe alguna diferencia. A pesar de que en la Figura 8 puede apreciarse una leve diferencia entre los grupos, el p-valor=0.2, nos indica que no existe una diferencia estadísticamente significativa. Al igual que en la situación de recaída con un p-valor 0.77 tanto

visualmente como estadísticamente podemos afirmar que no existe una diferencia estadísticamente significativa para estos dos tipos de cáncer sobre la situación de recaída.



**Figura 10:** Función de supervivencia tipo de tumor y sexo Situación 1 (Muerte)

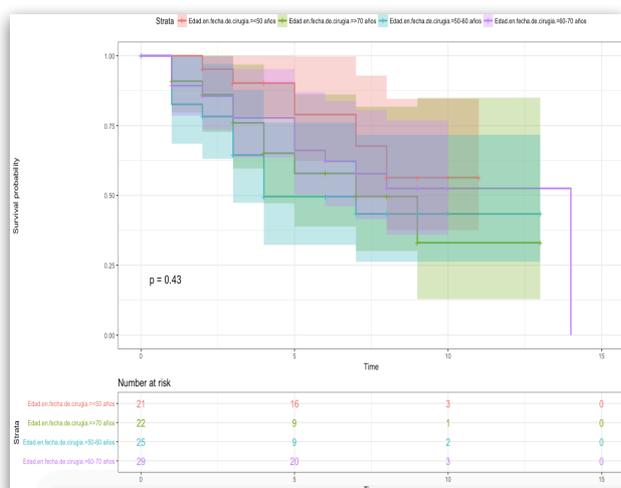


**Figura 11:** Función de supervivencia tipo de tumor y sexo Situación 2 (Recaída)

Como mencionamos en los primeros apartados de este capítulo, es interesante relacionar estas variables explicativas con el género. Comenzando con un diagnóstico estadístico, podemos observar que en todas las situaciones obtenemos un  $p\text{-valor} > 0.05$ , por tanto podemos afirmar que no existe una diferencia estadísticamente significativa entre curvas de supervivencia. Aún así, podemos obtener información de interés en cuanto la comparación por género. En la Figura 10 podemos observar que los pacientes diagnosticados con CCC tienen menos probabilidad de supervivencia que los paciente diagnosticados con CRP. A pesar de que para los hombres las curvas son muy parecidas, para las mujeres es interesante resaltar como la probabilidad supervivencia se iguala en los 8 años. Es decir, las mujeres diagnosticadas con *CRP* presentan un aumento de eventos que se iguala en probabilidad de supervivencia a las personas diagnosticadas con *CCC* en este tiempo.

Para la Figura 11 podemos observar que la probabilidad de recaída es muy similar para los dos tipos de diagnóstico tanto en hombres como en mujeres.

## Función de Supervivencia por rango de edad y sexo

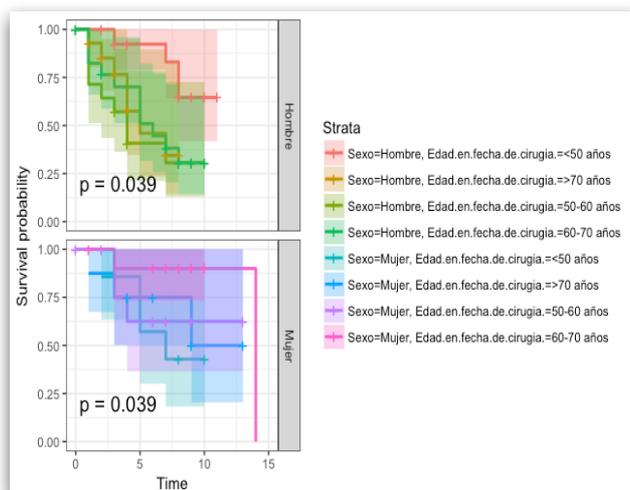


**Figura 12:** Función de supervivencia edad *Situación 1* (Muerte)

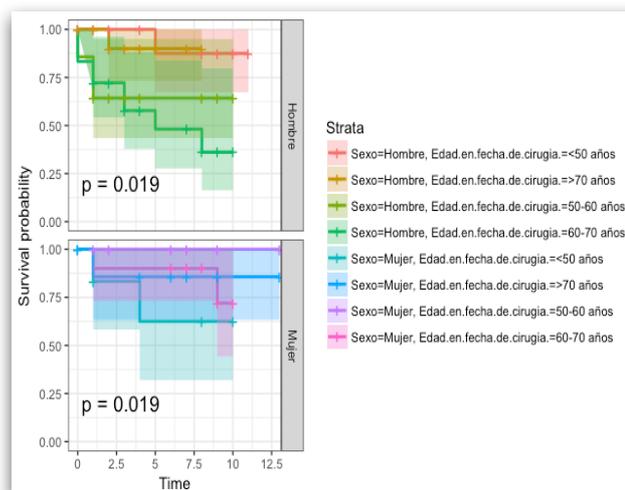


**Figura 13:** Función de supervivencia tipo edad *Situación 2* (Recaída)

En este apartado, analizamos la influencia de la edad sobre la supervivencia en pacientes con CCR. Para facilitar este análisis en las curvas de supervivencia, y guiándonos en estudios similares, se realizó una conversión de la variable (Edad) en rangos, como mencionábamos en el apartado de codificación de variables. Se analiza los valores obtenidos mediante el test de *Long-Rank*, para las dos situaciones. Puede verse que en ambas situación, el p-valor  $> 0.05$ , lo cual indica que no existe diferencia estadísticamente significativa entre curvas de supervivencia para los grupos de edad. A pesar de la no existencia de diferencia significativa, podemos analizar la influencia de la supervivencia y recaída en los diferentes rangos de edad. En la Figura 12 observamos que la probabilidad de supervivencia es menor para las personas entre 50-60 años, seguidos de las personas mayores de 70 años. Sin embargo, apreciamos que la supervivencia es mucho mejor para el rango de edad de menores de 50 años, algo que concuerda con la bibliografía. En la Figura 13 podemos destacar que la probabilidad de recaída en este tipo de cáncer es muy negativa para las personas entre 60-70 años.



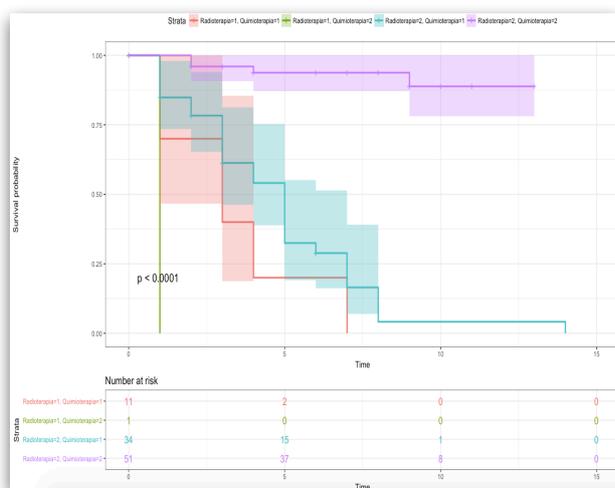
**Figura 14:** Función de supervivencia por edad y sexo *Situación 1 (Muerte)*



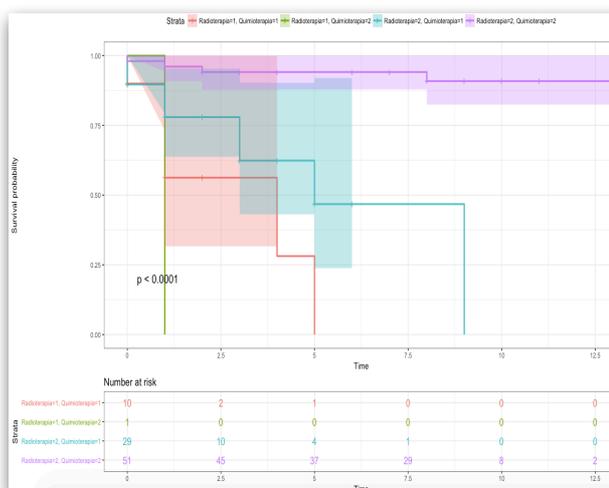
**Figura 15:** Función de supervivencia por edad y sexo *Situación 2 (Recaída)*

Ahora se contrasta las curvas de supervivencia pero separada por género. Observamos como mediante el test de Long-Rank obtenemos que en tanto en la situación de exitus por enfermedad (*Situación 1: P-valor=0.039*) como en recaída (*Situación 2: p-valor=0.019*), existe una diferencia estadísticamente significativa entre curvas de supervivencia. En la Figura 14 podemos observar como el CCR en hombres entre 50-60 años son los que presentan peor supervivencia. Para las mujeres es lo contrario el rango de edad, entre 50-60 años presentan una probabilidad de supervivencia más positiva. Pero la menor probabilidad de supervivencia la presentan los rangos de edad >70 años y <50 años. Aún así se destaca que durante los primeros 5 años después de la operación, la probabilidad de supervivencia es muy similar en los diferentes grupos de edad. En la Figura 15 podemos apreciar que los hombres presentan una destacable diferencia en las curvas. La probabilidad de recaída más alta se presenta en pacientes entre los 50-70 años. Para los extremos en este rango de edad, mayores de 70 y menores 50 años, tienen un mejor pronóstico en cuanto a probabilidad de recaída. La probabilidad de recaída en el grupo de las mujeres, destacamos que los grupos más vulnerables se encuentran en pacientes menores de 50 años y más favorable para los pacientes entre 50-70 años, caso totalmente contrario al grupo de hombres.

## **Función de Supervivencia radioterapia y(o) quimioterapia**

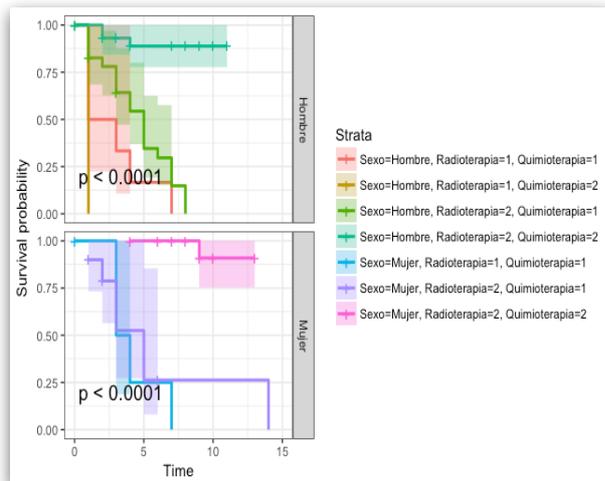


**Figura 16:** Función de supervivencia Radioterapia y(o) Quimioterapia *Situación 1 (Muerte)*

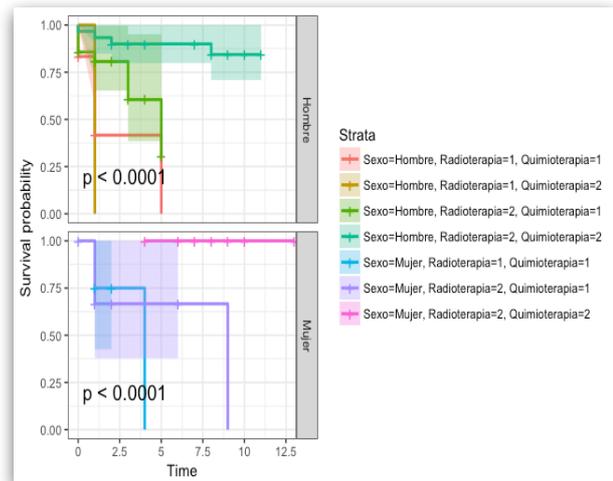


**Figura 17:** Función de supervivencia Radioterapia y(o) Quimioterapia *Situación 2 (Recaída)*

Se analiza ahora la influencia de los tratamientos de Radioterapia (1= Si ; 2= No) y Quimioterapia (1= Si ; 2= No) sobre las curvas de supervivencia en pacientes de CCR. Se propone el análisis de estas dos situaciones juntas debido a la interacción entre estas variables, ya que un mismo paciente puede haber recibido uno o ambos tratamientos. Estadísticamente podemos afirmar que existe una diferencia significativa en las curvas de supervivencia mediante el test de Long-Rank para las dos situaciones donde se obtuvo un p-valor = 0.001. En la Figura 16 podemos apreciar que la supervivencia disminuye en el grupo de pacientes que recibieron ambos tratamientos y los que sólo recibieron radioterapia. En cuanto a efectividad entre un tratamiento u otro, las curvas de supervivencia son mejores para los pacientes que solo recibieron quimioterapia, este caso se presenta para las dos situaciones. Sin embargo, los pacientes que no tuvieron ningún tratamiento, presentan una supervivencia más favorable. En la Figura 17 presenta un caso similar a la situación 1. La probabilidad de recaída es mayor para los pacientes que recibieron ambos tratamientos o solo radioterapia. Los pacientes que no recibieron ninguna terapia, tienen mejor pronóstico ante una probabilidad de recaída.



**Figura 18:** Función de supervivencia Radioterapia y(o) Quimioterapia y sexo Situación 1 (Muerte)



**Figura 19:** Función de supervivencia Radioterapia y(o) Quimioterapia y sexo Situación 2 (Recaída)

En este apartado analizamos la influencia de los tratamientos radioterapia y quimioterapia sobre las curvas de supervivencias y el género en pacientes con CCR. Para las dos situaciones se obtuvo un  $p$ -valor  $< 0.05$ , esto nos indica que existe una diferencia estadísticamente significativa en las curvas de supervivencia de estos grupos. En la Figura 18 podemos observar que en el grupo de hombres que recibió ambos tratamientos o sólo radioterapia, presentan una probabilidad de supervivencia menos favorable. La supervivencia más favorable la presentan los pacientes que no recibieron ninguno de los dos tratamientos. Aún así, cabe destacar que en comparación de tratamientos, la quimioterapia tiene una curva de supervivencia con pronóstico más favorable que la radioterapia. En las mujeres se presenta un caso muy similar, la supervivencia es más desfavorable para las pacientes que recibieron ambos tratamientos, seguida de las pacientes que solo recibieron Radioterapia. Pero la probabilidad de supervivencia más favorable la presenta el grupo que no recibió ningún tratamiento. En la Figura 19, podemos observar que el caso es exactamente el mismo que en la situación 1. La probabilidad de recaída es menos favorable para los pacientes que recibieron ambos tratamientos, tanto en hombres como en mujeres. La probabilidad de recaída tampoco favorece a quienes solo recibieron radioterapia. El caso favorable en cuanto al evento de recaída lo presentan los pacientes que no recibieron ningún tratamiento. También cabe destacar que si comparásemos las curvas entre tratamientos, la quimioterapia sigue siendo más favorable que la radioterapia en cuanto a recaída.

---

### 2.5.1 Estimación de modelo de Cox univariante y multivariante

Se analiza la influencia de las variables explicativas sobre el estudio. En el ámbito biomédico resulta útil aplicar modelos multivariantes que nos permitan examinar cómo la interacción de estas variables, trabajan juntas para influir en la velocidad de un evento en particular (muerte o recaída) en un momento en particular (López-Jiménez et al., 2005).

A continuación se evalúan en primer lugar un modelo univariante con tal de describir la influencia individual de cada covariable sobre las dos situaciones. Consiguientemente ajustaremos un modelo multivariante mediante la estimación del modelo de riesgos proporcionales de Cox.

### 2.5.2 Estimación del modelo de Cox univariante

#### Situación 1

	beta	HR (95% CI for HR)	wald.test	p.value
Sexo	-0.73	0.48(0.24-0.96)	4.3	0.039
Quimioterapia	-3.1	0.046(0.018-0.12)	38	6.7e-10
Radioterapia	-1.8	0.17(0.083-0.35)	23	1.3e-06
Edad.en.fecha.de.cirugia.	0.0073	1 (0.98-1)	0.32	0.57

Los resultados para este análisis univariante, se recopilan en el data.frame() anterior con la información del coeficiente de regresión beta, los tamaños del efecto (dados como cocientes de riesgo) y la significación estadística para cada una de las variable en relación con la supervivencia.

Podemos observar que las variables sexo, quimioterapia y radioterapia tienen un p-valor < 0.05, estadísticamente significativos. La edad es la única covariable con coeficientes beta positivo, sin embargo, no es estadísticamente significativa. Los coeficientes beta negativos, son asociados como indicador de menores tasas de supervivencia para estas variables.

---

## Situación 2

	beta	HR (95% CI for HR)	wald.test	p.value
Sexo	-0.82	0.44 (0.16-1.2)	2.5	0.11
Quimioterapia	-2.4	0.094(0.031-0.28)	18	2.5e-05
Radioterapia	-2	0.14 (0.052-0.36)	16	5.9e-05
Edad.en.fecha.de.cirugia.	0.0018	1 (0.97-1)	0.01	0.92

Para la situación de recaída, las variables de sexo y edad presentan p-valor  $> 0.05$ , unos coeficientes que no son estadísticamente significativos. Pero, las variables radioterapia y quimioterapia siguen presentado coeficientes altamente significativos. En cuanto a los coeficientes beta, las variables sexo, quimioterapia y radioterapia presentan valores negativos, lo cual indica que para estas variables hay mejores pronósticos ante una situación de recaída.

### 2.5.2 Estimación de modelo de Cox multivariante

Inicialmente creamos un modelo teniendo en cuenta todas las variables y analizaremos la situación cuando se consideran juntas y su influencia en la supervivencia. Luego reduciremos el espacio de las variables explicativas hasta obtener un modelo significativo que pueda ayudarnos a explicar la influencia de estas variables con la supervivencia en pacientes con CCR.

#### **Modelo de Cox ajustado para la situación 1**

En primer lugar observamos una descripción de las variables que utilizaremos para el modelo:

```
'data.frame':      99 obs. of  11 variables:
 $ ID                : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Sexo : Factor w/ 2 levels "Hombre","Mujer": 2 2 2 1 2 1 2 1 1 1 ...
 $ Edad.en.fecha.de.cirugia.: num  67 59 74 54 52 38 47 77 73 64 ...
 $ Tipo.de.tumor : Factor w/ 10 levels "CCC","CCC + CRS",...: 1 6 1 1 1 1 6 1...
 $ Estadio : Factor w/ 51 levels "Estadio IV","I",...: 20 2 8 NA NA 5 6 25 ...
 $ Evento_1      : num  1 0 0 1 0 0 1 0 1 1 ...
 $ Evento_2      : num  1 0 0 1 0 0 1 1 1 0 ...
 $ Quimioterapia : num  1 2 2 2 2 2 1 1 1 2 ...
 $ Radioterapia  : num  2 2 2 2 2 2 1 2 2 2 ...
 $ Tiempo_Rec    : num  3403 4605 4663 492 2469 ...
 $ Tiempo_Sup    : num  5055 4605 4663 759 2469 ...
```

En este primer esbozo del modelo, usaremos las variables “Edad” y “Tipo de Tumor” sin la modificación que hemos propuesto en epígrafes anteriores. Usaremos el método “Backward” para la selección de variables explicativas :

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ Sexo + Edad.en.fecha.de.cirugia. +
      Tipo.de.tumor + Quimioterapia + Radioterapia, data = data_CCR)
```

	coef	exp(coef)	se(coef)	z	p
SexoMujer	-3.00e-01	7.41e-01	3.65e-01	-0.82	0.41
Edad.en.fecha.de.cirugia.	1.99e-02	1.02e+00	1.51e-02	1.32	0.19
Tipo.de.tumorCCC + CRS	7.21e-01	2.06e+00	7.21e-01	1.00	0.32
Tipo.de.tumorCCR	9.75e-04	1.00e+00	1.64e+00	0.00	1.00
Tipo.de.tumorCRC + CRS	1.25e-01	1.13e+00	1.19e+00	0.11	0.92
Tipo.de.tumorCRI	NA	NA	0.00e+00	NA	NA
Tipo.de.tumorCRP	-2.36e-01	7.90e-01	4.50e-01	-0.53	0.60
Tipo.de.tumorM + A + CCC	-5.08e-02	9.50e-01	8.03e-01	-0.06	0.95
Tipo.de.tumorM + A + CRP	-1.73e+01	3.06e-08	1.02e+04	0.00	1.00
Tipo.de.tumorM + CCC	-8.41e+00	2.23e-04	1.08e+02	-0.08	0.94
Tipo.de.tumorM + CRP	1.50e-01	1.16e+00	1.09e+00	0.14	0.89
Quimioterapia	-1.96e+00	1.41e-01	4.09e-01	-4.79	1.7e-06
Radioterapia	-6.04e-01	5.47e-01	5.87e-01	-1.03	0.30

Likelihood ratio test=37.4 on 12 df, p=0.000196

n= 97, number of events= 43

(2 observations deleted due to missingness)

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ Sexo + Edad.en.fecha.de.cirugia. +
      Tipo.de.tumor + Quimioterapia + Radioterapia, data = data_CCR)
```

n= 97, number of events= 43

(2 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )
SexoMujer	-3.003e-01	7.406e-01	3.645e-01	-0.824	0.410
Edad.en.fecha.de.cirugia.	1.990e-02	1.020e+00	1.508e-02	1.320	0.187
Tipo.de.tumorCCC + CRS	7.209e-01	2.056e+00	7.209e-01	1.000	0.317
Tipo.de.tumorCCR	9.748e-04	1.001e+00	1.637e+00	0.001	1.000
Tipo.de.tumorCRC + CRS	1.248e-01	1.133e+00	1.188e+00	0.105	0.916
Tipo.de.tumorCRI	NA	NA	0.000e+00	NA	NA
Tipo.de.tumorCRP	-2.361e-01	7.897e-01	4.496e-01	-0.525	0.600
Tipo.de.tumorM + A + CCC	-5.083e-02	9.504e-01	8.034e-01	-0.063	0.950

---

```

Tipo.de.tumorM + A + CRP  -1.730e+01  3.060e-08  1.024e+04 -0.002  0.999
Tipo.de.tumorM + CCC      -8.410e+00  2.226e-04  1.077e+02 -0.078  0.938
Tipo.de.tumorM + CRP      1.496e-01  1.161e+00  1.090e+00  0.137  0.891
Quimioterapia             -1.956e+00  1.414e-01  4.088e-01 -4.785  1.71e-06 ***
Radioterapia              -6.038e-01  5.467e-01  5.866e-01 -1.029  0.303

```

---

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

```

                                exp(coef) exp(-coef) lower .95 upper .95
SexoMujer                       7.406e-01  1.350e+00  3.625e-01  1.513e+00
Edad.en.fecha.de.cirugia. 1.020e+00  9.803e-01  9.904e-01  1.051e+00
Tipo.de.tumorCCC + CRS          2.056e+00  4.863e-01  5.006e-01  8.447e+00
Tipo.de.tumorCCR                1.001e+00  9.990e-01  4.043e-02  2.478e+01
Tipo.de.tumorCRC + CRS          1.133e+00  8.826e-01  1.103e-01  1.164e+01
Tipo.de.tumorCRI                 NA          NA          NA          NA
Tipo.de.tumorCRP                7.897e-01  1.266e+00  3.271e-01  1.906e+00
Tipo.de.tumorM + A + CCC        9.504e-01  1.052e+00  1.968e-01  4.589e+00
Tipo.de.tumorM + A + CRP        3.060e-08  3.268e+07  0.000e+00      Inf
Tipo.de.tumorM + CCC            2.226e-04  4.493e+03  5.169e-96  9.581e+87
Tipo.de.tumorM + CRP            1.161e+00  8.610e-01  1.371e-01  9.841e+00
Quimioterapia                   1.414e-01  7.072e+00  6.346e-02  3.151e-01
Radioterapia                    5.467e-01  1.829e+00  1.732e-01  1.726e+00

```

Concordance= 0.788 (se = 0.051 )

Rsquare= 0.32 (max possible= 0.971 )

Likelihood ratio test= 37.35 on 12 df, p=0.0001957

Wald test = 37.39 on 12 df, p=0.0001931

Score (logrank) test = 105.8 on 12 df, p=0

El valor de la "Concordance= 0.788" funciona como el coeficiente de determinación en regresión un ( $R^2$ ), indicando la relación del modelo con las covariables explicativas. Podemos observar que este valor no es bastante bueno, aunque el modelo es significativo por los test de Likelihood, Wald, Score (logrank). La única variable explicativa con significación es la Quimioterapia. Consideramos una reducción de las variables utilizando las más significativas.

---

Para este segundo modelo se utilizan la variable codificada de rango de edad:

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia, data = data_CCR)
```

	coef	exp(coef)	se(coef)	z	p
Edad.en.fecha.de.cirugia.>70 años	1.3061	3.6917	0.5060	2.58	0.0099
Edad.en.fecha.de.cirugia.50-60 años	0.9901	2.6914	0.4750	2.08	0.0371
Edad.en.fecha.de.cirugia.60-70 años	0.5798	1.7857	0.4982	1.16	0.2445
Quimioterapia	-3.1095	0.0446	0.5228	-5.95	2.7e-09
Radioterapia	-0.8733	0.4176	0.4093	-2.13	0.0329

Likelihood ratio test=75.9 on 5 df, p=6e-15

n= 97, number of events= 43

(2 observations deleted due to missingness)

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ Edad.en.fecha.de.cirugia.+Quimioterapia + Radioterapia, data = data_CCR)
```

n= 97, number of events= 43

(2 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )
Edad.en.fecha.de.cirugia.>70 años	1.30608	3.69168	0.50603	2.581	0.00985 **
Edad.en.fecha.de.cirugia.50-60 años	0.99007	2.69144	0.47501	2.084	0.03713 *
Edad.en.fecha.de.cirugia.60-70 años	0.57981	1.78569	0.49816	1.164	0.24446
Quimioterapia	-3.10951	0.04462	0.52283	-5.947	2.72e-09 ***
Radioterapia	-0.87331	0.41757	0.40932	-2.134	0.03288 *

---

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

	exp(coef)	exp(-coef)	lower .95	upper .95
Edad.en.fecha.de.cirugia.>70 años	3.69168	0.2709	1.36928	9.9531
Edad.en.fecha.de.cirugia.50-60 años	2.69144	0.3715	1.06086	6.8283
Edad.en.fecha.de.cirugia.60-70 años	1.78569	0.5600	0.67263	4.7406
Quimioterapia	0.04462	22.4101	0.01601	0.1243
Radioterapia	0.41757	2.3948	0.18720	0.9314

Concordance= 0.848 (se = 0.051 )

Rsquare= 0.543 (max possible= 0.971 )

---

Likelihood ratio test= 75.93 on 5 df, p=5.995e-15  
 Wald test = 47.49 on 5 df, p=4.507e-09  
 Score (logrank) test = 80.8 on 5 df, p=5.551e-16

Este modelo 2 es mucho más significativo en cuanto *concordance=0.848*. También podemos observar que se presentan variables estadísticamente significativas, además tiene un p-valor significativo muy similar mediante los tres test Likelihood, Wald, Score (logrank).

Verificamos la hipótesis de riesgos proporcionales mediante la función *cox.zph()* la cual usa los residuos de *Shoenfeld* y se basa en el test de *Kolmogorov*.

	rho	chisq	p
Edad.en.fecha.de.cirugia.>70 años	-0.0332	0.0453	0.8315
Edad.en.fecha.de.cirugia.50-60 años	-0.1634	1.0247	0.3114
Edad.en.fecha.de.cirugia.60-70 años	-0.2311	2.5992	0.1069
Quimioterapia	-0.3257	5.3129	0.0212
Radioterapia	0.0554	0.1525	0.6962
GLOBAL	NA	8.4678	0.1323

Para comprobar la hipótesis de riesgos proporcionales observamos los p-valor. En la mayoría de las variables los *p-valores* > 0.05 lo cual no tenemos motivos para rechazar la hipótesis de riesgos proporcionales. A pesar de que se presenta esta situación, es necesario pensar en una alternativa a este modelo porque la variable *Quimioterapia* tiene un p-valor < 0.05 y lo cual supone rechazar la hipótesis de que se cumpla la proporcionalidad de riesgos.

Para solucionar el problema del modelo cuando no cumple con el supuesto de proporcionalidad de los riesgos, la teoría nos ofrece dos opciones; la primera opción es evaluar un modelo de Cox con variables dependientes del tiempo. La segunda opción es utilizar un modelo de Cox estratificado, este modelo permite corregir el modelo cuando una de las covariables no cumple el supuesto de riesgos proporcionales (Pena, Rafael Eduardo Borges 2005), el cual es nuestro caso. Aplicaremos entonces la función *Strata()* para la covariable *Quimioterapia* y evaluamos de nuevo el modelo.

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ strata(Quimioterapia) + Radioterapia + Edad.en.fecha.de.cirugia., data = data_CCR)
```

	coef	exp(coef)	se(coef)	z	p
Radioterapia	-0.964	0.381	0.416	-2.32	0.0206
Edad.en.fecha.de.cirugia.>70 años	1.411	4.101	0.516	2.73	0.0063
Edad.en.fecha.de.cirugia.50-60 años	1.128	3.088	0.484	2.33	0.0198
Edad.en.fecha.de.cirugia.60-70 años	0.649	1.913	0.500	1.30	0.1948

Likelihood ratio test=12.9 on 4 df, p=0.0119

n= 97, number of events= 43

(2 observations deleted due to missingness)

Call:

```
coxph(formula = Surv(Tiempo_Sup, Evento_2) ~ strata(Quimioterapia) + Radioterapia + Edad.en.fecha.de.cirugia., data = data_CCR)
```

n= 97, number of events= 43

(2 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )
Radioterapia	-0.9637	0.3815	0.4162	-2.315	0.02059 *
Edad.en.fecha.de.cirugia.>70 años	1.4112	4.1010	0.5163	2.733	0.00627 **
Edad.en.fecha.de.cirugia.50-60 años	1.1276	3.0883	0.4838	2.331	0.01978 *
Edad.en.fecha.de.cirugia.60-70 años	0.6487	1.9131	0.5004	1.297	0.19480

---

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

	exp(coef)	exp(-coef)	lower .95	upper .95
Radioterapia	0.3815	2.6214	0.1687	0.8625
Edad.en.fecha.de.cirugia.>70 años	4.1010	0.2438	1.4908	11.2818
Edad.en.fecha.de.cirugia.50-60 años	3.0883	0.3238	1.1964	7.9719
Edad.en.fecha.de.cirugia.60-70 años	1.9131	0.5227	0.7175	5.1009

Concordance= 0.693 (se = 0.084 )

Rsquare= 0.124 (max possible= 0.923 )

Likelihood ratio test= 12.87 on 4 df, p=0.0119

Wald test = 12.67 on 4 df, p=0.013

Score (logrank) test = 13.45 on 4 df, p=0.009279

	rho	chisq	p
Radioterapia	0.0332	0.05723	0.811
Edad.en.fecha.de.cirugia.>70 años	-0.0139	0.00829	0.927
Edad.en.fecha.de.cirugia.50-60 años	-0.1379	0.76559	0.382
Edad.en.fecha.de.cirugia.60-70 años	-0.2059	2.10322	0.147
GLOBAL	NA	3.34515	0.502

Con este modelo, se ve disminuida la concordance = 0.693. Aún así, el p-valor para las tres pruebas generales es significativo, lo que indica que el modelo es significativo. Utilizando la función `cox.zph()` observamos que este modelo cumple con la condición de riesgos proporcionales.

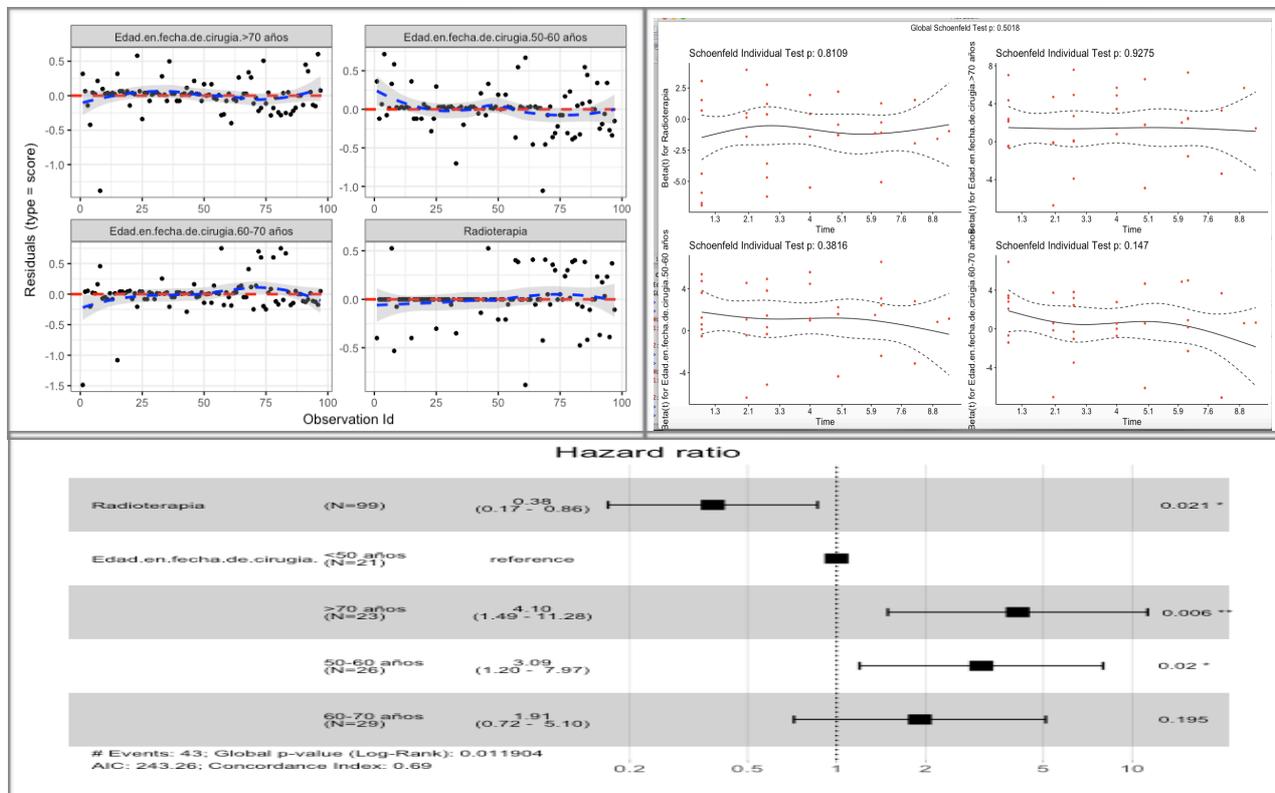


Figura 20: Análisis de residuos *Situación 1*

La Figura 20 muestra los diagramas de distribución de los residuales parciales contra los rangos de tiempo de supervivencia. No se observa ningún patrón de distribución de los residuos que sugiera falta de ajuste de los datos al modelo. Los residuos son independientes del tiempo de supervivencia.

El diagrama de bosques muestra la relación de riesgo (HR) y los intervalos de confianza del 95%. La línea vertical en 1 se conoce como la línea de efecto nulo para HR, indica la igualdad en la probabilidad de muerte dentro de los grupos de la variable en el siguiente intervalo de tiempo. Un  $HR > 1$  Indica mayor factor de riesgo junto con coeficientes beta positivos y una  $HR < 1$  indica una reducción del riesgo, por tanto, un aumento en la probabilidad de supervivencia correspondiéndose con factores de protección y coeficientes beta negativos (Molina, M 2015). Podemos observar que la variable radioterapia se encuentra por debajo de la línea del efecto nulo, las personas que recibieron radioterapia contemplan mejor probabilidad de supervivencia frente a las que no recibieron dicho tratamiento. Sin embargo, las demás variables presentan un  $HR > 1$  indicando mayor factor de riesgo dentro de los grupos.

### **Modelo de Cox ajustado para la situación 2**

Ajustamos ahora un modelo para la situación de recaída. Igualmente que el modelo anterior, evaluamos un primer modelo con todas las variables explicativas.

Call:

```
coxph(formula = Surv(Tiempo_Rec, Evento_1) ~ Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia + Tipo.de.tumor, data = data_CCR)
```

	coef	exp(coef)	se(coef)	z	p
Edad.en.fecha.de.cirugia.	2.29e-02	1.02e+00	2.25e-02	1.02	0.30792
Quimioterapia	-2.36e+00	9.40e-02	6.26e-01	-3.78	0.00016
Radioterapia	-1.68e+00	1.87e-01	6.37e-01	-2.63	0.00851
Tipo.de.tumorCCC + CRS	-1.52e-01	8.59e-01	1.13e+00	-0.14	0.89248
Tipo.de.tumorCCR	-1.51e+01	2.68e-07	5.35e+03	0.00	0.99774
Tipo.de.tumorCRC + CRS	-2.20e+01	2.85e-10	4.59e+04	0.00	0.99962
Tipo.de.tumorCRI	NA	NA	0.00e+00	NA	NA
Tipo.de.tumorCRP	-7.20e-02	9.30e-01	5.94e-01	-0.12	0.90341
Tipo.de.tumorM + A + CCC	-2.03e+01	1.51e-09	3.81e+04	0.00	0.99957
Tipo.de.tumorM + A + CRP	NA	NA	0.00e+00	NA	NA
Tipo.de.tumorM + CCC	-6.72e-01	5.11e-01	1.17e+00	-0.58	0.56404

---

Tipo.de.tumorM + CRP                    NA            NA 0.00e+00    NA        NA

Likelihood ratio test=45.5 on 9 df, p=7.49e-07

n= 91, number of events= 21

(8 observations deleted due to missingness)

Call:

```
coxph(formula = Surv(Tiempo_Rec, Evento_1) ~ Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia + Tipo.de.tumor, data = data_CCR)
```

n= 91, number of events= 21

(8 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )	
Edad.en.fecha.de.cirugia.	2.292e-02	1.023e+00	2.248e-02	1.020	0.307916	
Quimioterapia	-2.364e+00	9.404e-02	6.255e-01	-3.779	0.000157	***
Radioterapia	-1.676e+00	1.870e-01	6.372e-01	-2.631	0.008512	**
Tipo.de.tumorCCC + CRS	-1.521e-01	8.589e-01	1.125e+00	-0.135	0.892476	
Tipo.de.tumorCCR	-1.513e+01	2.680e-07	5.354e+03	-0.003	0.997745	
Tipo.de.tumorCRC + CRS	-2.198e+01	2.849e-10	4.585e+04	0.000	0.999618	
Tipo.de.tumorCRI	NA	NA	0.000e+00	NA	NA	
Tipo.de.tumorCRP	-7.203e-02	9.305e-01	5.936e-01	-0.121	0.903413	
Tipo.de.tumorM + A + CCC	-2.031e+01	1.512e-09	3.810e+04	-0.001	0.999575	
Tipo.de.tumorM + A + CRP	NA	NA	0.000e+00	NA	NA	
Tipo.de.tumorM + CCC	-6.722e-01	5.106e-01	1.165e+00	-0.577	0.564043	
Tipo.de.tumorM + CRP	NA	NA	0.000e+00	NA	NA	

---

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

	exp(coef)	exp(-coef)	lower .95	upper .95
Edad.en.fecha.de.cirugia.	1.023e+00	9.773e-01	0.97908	1.0693
Quimioterapia	9.404e-02	1.063e+01	0.02760	0.3204
Radioterapia	1.870e-01	5.346e+00	0.05365	0.6521
Tipo.de.tumorCCC + CRS	8.589e-01	1.164e+00	0.09463	7.7957
Tipo.de.tumorCCR	2.680e-07	3.732e+06	0.00000	Inf
Tipo.de.tumorCRC + CRS	2.849e-10	3.510e+09	0.00000	Inf
Tipo.de.tumorCRI	NA	NA	NA	NA
Tipo.de.tumorCRP	9.305e-01	1.075e+00	0.29070	2.9784
Tipo.de.tumorM + A + CCC	1.512e-09	6.616e+08	0.00000	Inf
Tipo.de.tumorM + A + CRP	NA	NA	NA	NA

Tipo.de.tumorM + CCC	5.106e-01	1.959e+00	0.05201	5.0121
Tipo.de.tumorM + CRP	NA	NA	NA	NA

Concordance= 0.839 (se = 0.067 )

Rsquare= 0.393 (max possible= 0.849 )

Likelihood ratio test= 45.49 on 9 df, p=7.492e-07

Wald test = 28.16 on 9 df, p=0.0008965

Score (logrank) test = 139.5 on 9 df, p=0

Podemos observar que el modelo tiene una *concordance*=0.839, es un buen valor para describir la situación de recaída. Pero solo dos variables son significativas, lo cual nos lleva a replantear y reducir el numero de variables.

Call:

```
coxph(formula = Surv(Tiempo_Rec, Evento_1) ~ (Edad.en.fecha.de.cirugia. == "60-70 años") +
Quimioterapia + Radioterapia + (Tipo.de.tumor == "CRI"), data = data_CCR)
```

	coef	exp(coef)	se(coef)	z	p
Edad.en.fecha.de.cirugia. == "60-70 años"TRUE	1.079	2.941	0.518	2.08	0.03743
Quimioterapia	-2.174	0.114	0.641	-3.39	0.00070
Radioterapia	-1.580	0.206	0.585	-2.70	0.00691
Tipo.de.tumor == "CRI"TRUE	5.020	151.402	1.295	3.88	0.00011

Likelihood ratio test=39.3 on 4 df, p=5.93e-08

n= 91, number of events= 21

(8 observations deleted due to missingness)

Call:

```
coxph(formula = Surv(Tiempo_Rec, Evento_1) ~ (Edad.en.fecha.de.cirugia. == "60-70 años") +
Quimioterapia + Radioterapia + (Tipo.de.tumor == "CRI"), data = data_CCR)
```

n= 91, number of events= 21

(8 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z )
Edad.en.fecha.de.cirugia. == "60-70 años"TRUE	1.0786	2.9406	0.5183	2.081	0.037434 *
Quimioterapia	-2.1740	0.1137	0.6411	-3.391	0.000696 ***
Radioterapia	-1.5801	0.2060	0.5849	-2.701	0.006907 **

---

```

Tipo.de.tumor == "CRI"TRUE          5.0199  151.4017   1.2946  3.878 0.000105 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                exp(coef) exp(-coef) lower .95 upper .95
Edad.en.fecha.de.cirugia. == "60-70 años"TRUE  2.9406  0.340072  1.06474  8.1211
Quimioterapia                0.1137  8.793018  0.03237  0.3995
Radioterapia                 0.2060  4.855215  0.06545  0.6482
Tipo.de.tumor == "CRI"TRUE    151.4017  0.006605  11.97198 1914.6772

Concordance= 0.821  (se = 0.067 )

Rsquare= 0.351  (max possible= 0.853 )

Likelihood ratio test= 39.34  on 4 df,  p=5.929e-08

Wald test          = 33.78  on 4 df,  p=8.25e-07

Score (logrank) test = 68.1  on 4 df,  p=5.707e-14

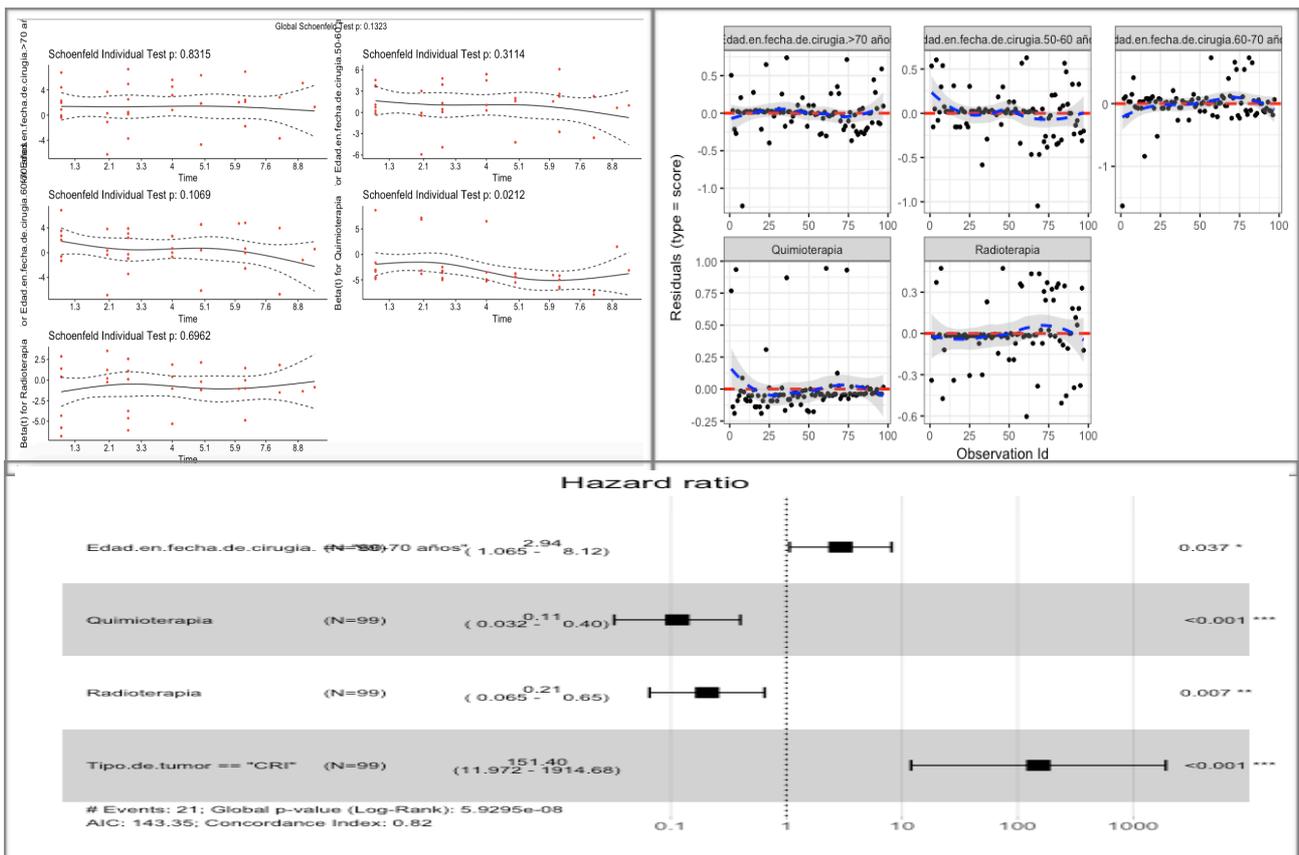
```

Este modelo se ajustó con la variable (rango de edad 60-70 años). tenemos una concordance = 0.821, es una buena correlación y se ajustan las variables más significativas. Además el modelo es significativo por los tres test Likelihood, Wald, Score (logrank)

Comprobamos la hipótesis de riesgos proporcionales del modelo:

	rho	chisq	p
Edad.en.fecha.de.cirugia. == "60-70 años"TRUE	0.189	0.614	0.433
Quimioterapia	-0.148	0.563	0.453
Radioterapia	-0.193	0.855	0.355
Tipo.de.tumor == "CRI"TRUE	0.083	0.141	0.707
GLOBAL	NA	2.174	0.704

Los p-valor > 0.05, indican que el modelo cumple con la hipótesis de proporcionalidad de riesgos.



**Figura 21:** Análisis de residuos *Situación 2*

La Figura 21 comprueba gráficamente que no se viola la hipótesis de riesgos proporcionales, tampoco puede apreciarse ningún patrón de distribución de los residuos que demuestre lo contrario. En la recta  $\beta(t)=0$ , los residuos serían independientes del tiempo de supervivencia.

El diagrama de bosques, presenta un HR < 1 para las variables de radioterapia y quimioterapia, lo cual se asocia a menores factores de riesgo para los pacientes que recibieron dichos tratamientos en cuenta a recaída. Para la variable de edad pasar del rango de 60-70 años, presenta un aumento de riesgo y finalmente para los pacientes de covariable (tipo de tumor no clasificado), presentan un aumento de riesgo en comparación con los demás tipos de tumores.

---

### 2.4.3 *Estimación de modelo paramétrico*

A diferencia de un modelo de riesgos proporcionales, un modelo paramétrico asume que el tiempo de supervivencia acelera o desacelera según un factor constante al comparar los niveles de las covariables. Por lo tanto, un modelo paramétrico asume que el tiempo de supervivencia sigue una distribución de probabilidad concreta.

Se prevé ajustar un modelo paramétrico a nuestras dos situaciones. Pero, debemos resaltar que en la primera estimación, obtuvimos un mensaje “*Invalid survival times for this distribution*”. Esto es debido a que nuestra dos situaciones las variables respuestas presentan tiempos iguales a cero. Este error es causado debido a que las distribuciones necesitan tiempos mayores que cero, de no ser así, el resultado dará un estimador indefinido, lo cual arroja este error.

Una posible solución es eliminar las filas cuyos resultados sean ceros y ajustamos los modelos Exponencial y Weibull.

## Modelo Exponencial

### Situación 1

Call:

```
flexsurvreg(formula = Surv(Tiempo_Sup, Evento_2) ~
Sexo + Edad.en.fecha.de.cirugia. +
```

```
Quimioterapia + Radioterapia, data = data_CCR,
dist = "exponential")
```

Estimates:

	data	mean	est	L95%	U95%	se	exp(est)
rate	NA	6.9157	1.4162	33.7726	5.5957	NA	
SexoMujer	0.3723	-0.4518	-1.1421	0.2386	0.3522	0.6365	
Edad >70años	0.2340	0.9049	-0.0489	1.8587	0.4866	2.4717	
Edad 50-60años	0.2447	0.6920	-0.2068	1.5908	0.4586	1.9977	
Edad 60-70años	0.2979	0.4794	-0.4639	1.4227	0.4813	1.6151	
Quimioterapia	1.5426	-2.5706	-3.5389	-1.6024	0.4940	0.0765	
Radioterapia	1.8830	-0.7102	-1.4933	0.0730	0.3996	0.4916	

	L95%	U95%
rate	NA	NA
SexoMujer	0.3191	1.2695
Edad.en.fecha.de.cirugia.>70 años	0.9523	6.4152
Edad.en.fecha.de.cirugia.50-60 años	0.8132	4.9074
Edad.en.fecha.de.cirugia.60-70 años	0.6288	4.1482
Quimioterapia	0.0290	0.2014
Radioterapia	0.2246	1.0757

N = 94, Events: 43, Censored: 51

Total time at risk: 535

Log-likelihood = -118.3348, df = 7

AIC = 250.6695

	Value	Std. Error	z	p	
(Intercept)	= 0.145	-1.93	0.809	-2.390	1.68e-02
SexoMujer	= 1.571	0.452	0.352	1.283	2.00e-01
Edad >70 años	= 0.405	-0.905	0.487	-1.860	6.30e-02
Edad 50-60 años	= 0.501	-0.692	0.459	-1.509	1.31e-01
Edad 60-70 años	= 0.619	-0.479	0.481	-0.996	3.19e-01
Quimioterapia	= 13.074	2.571	0.494	5.204	1.96e-07
Radioterapia	= 2.034	0.710	0.400	1.777	7.55e-02

### Situación 2

Call:

```
flexsurvreg(formula = Surv(Tiempo_Rec, Evento_1) ~
Sexo + Edad.en.fecha.de.cirugia. +
```

```
Quimioterapia + Radioterapia, data = data_CCR,
dist = "exponentia")
```

Estimates:

	data	mean	est	L95%	U95%	se	exp(est)
rate	NA	15.3114	2.0687	113.3286	15.6374	NA	
SexoMujer	0.3780	-0.4981	-1.5899	0.5937	0.5570	0.6077	
Edad >70años	0.2195	0.5368	-1.2913	2.3650	0.9327	1.7106	
Edad 50-60años	0.2439	0.5961	-1.0402	2.2323	0.8348	1.8149	
Edad 60-70años	0.3049	1.5167	0.0291	3.0042	0.7590	4.5570	
Quimioterapia	1.6220	-1.8307	-3.1050	-0.5565	0.6501	0.1603	
Radioterapia	1.8902	-2.1094	-3.3798	-0.8391	0.6481	0.1213	

	L95%	U95%
rate	NA	NA
SexoMujer	0.2040	1.8107
Edad.en.fecha.de.cirugia.>70 años	0.2749	10.6438
Edad.en.fecha.de.cirugia.50-60 años	0.3534	9.3209
Edad.en.fecha.de.cirugia.60-70 años	1.0296	20.1700
Quimioterapia	0.0448	0.5732
Radioterapia	0.0341	0.4321

N = 82, Events: 16, Censored: 66

Total time at risk: 437

Log-likelihood = -51.29772, df = 7

AIC = 116.5954

	Value	Std. Error	z	p	
(Intercept)	= 0.045	-2.729	1.021	-2.672	0.00755
SexoMujer	= 1.685	0.498	0.557	0.894	0.37124
Edad >70 años	= 0.459	-0.537	0.933	-0.576	0.56492
Edad 50-60 años	= 0.430	-0.596	0.835	-0.714	0.47523
Edad 60-70 años	= 0.615	-1.517	0.759	-1.998	0.04568
Quimioterapia	= 20.001	1.831	0.650	2.816	0.00486
Radioterapia	= 2.330	2.109	0.648	3.255	0.00114

---

Evaluamos las dos situaciones mediante un modelo exponencial. Como mencionamos anteriormente la función de riesgo del modelo exponencial es la forma más sencilla de la distribución de Weibull. También indicamos que esta distribución tiene la particularidad de un riesgo es constante a lo largo del tiempo.

Es necesario tener en cuenta que un coeficiente positivo (negativo) indica que el efecto de la variable aumenta (disminuye) el tiempo hasta la muerte. Para la situación 1, En relación con las variables significativas. En cuanto al sexo, el tiempo de fallo aumenta 1.571 para las mujeres respecto a los hombres. Para la variable quimioterapia el tiempo de fallo aumenta 13.074 para los que recibieron este tratamiento en comparación a los que no lo recibieron. Para la radioterapia, el tiempo hasta el fallo es de 2.034 para los que la recibieron el tratamiento en comparación a los que no lo recibieron. Todos los p-valores  $< 0.05$ , indican que todas las variables tienen un efecto estadísticamente significativo en el factor de aceleración del tiempo de fallo. Para la situación 2, las variables estadísticamente significativas resaltamos el tiempo de fallo para los pacientes que recibieron radioterapia aumenta 2.330 en comparación con los que no recibieron el tratamiento. Para los pacientes que recibieron quimioterapia esta aceleración de fallo aumenta 20.001 con respecto a los que no la recibieron. Al límite con un p-valor estadísticamente significativo se encuentran los pacientes entre 60-70 años, los cuales tienen un tiempo de fallo de 0.615 con respecto a los pacientes que no se encuentran en este rango de edad.

## Modelo Weibull

### Situación 1

Call:

```
survreg(formula = Surv(Tiempo_Sup, Evento_2) ~ Sexo +  
Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia, data = data_CCR,  
dist = "weibull")
```

Coefficients:

(Intercept)	SexoMujer
-0.7941310	0.5148838
Edad.en.fecha.de.cirugia.>70 años	Edad.en.fecha.de.cirugia.50-60 años
-0.8561106	-0.8535969
Edad.en.fecha.de.cirugia.60-70 años	Quimioterapia
-0.5673112	1.2881994
Radioterapia	
0.9073326	

Scale= 0.497406

Loglik(model)= -88.6 Loglik(intercept only)= -124

Chisq= 70.88 on 6 degrees of freedom, p= 2.7e-13

n=82 (6 observations deleted due to missingness)

Call:

```
survreg(formula = Surv(Tiempo_Sup, Evento_2) ~ Sexo +  
Edad.en.fecha.de.cirugia. +  
Quimioterapia + Radioterapia, data = data_CCR, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.794	0.481	-1.65	9.89e-02
SexoMujer	0.515	0.215	2.39	1.68e-02
Edad.en.fecha.de.cirugia.>70 años	-0.856	0.303	-2.83	4.70e-03
Edad.en.fecha.de.cirugia.50-60 años	-0.854	0.282	-3.02	2.50e-03
Edad.en.fecha.de.cirugia.60-70 años	-0.567	0.307	-1.85	6.45e-02
Quimioterapia	1.288	0.298	4.32	1.54e-05
Radioterapia	0.907	0.263	3.45	5.56e-04
Log(scale)	-0.698	0.139	-5.01	5.47e-07

Scale= 0.497

Weibull distribution

Loglik(model)= -88.6 Loglik(intercept only)= -124

Chisq= 70.88 on 6 degrees of freedom, p=  
2.7e-13

Number of Newton-Raphson Iterations: 7

n=82 (6 observations deleted due to missingness)

### Situación 2

Call:

```
survreg(formula = Surv(Tiempo_Rec, Evento_1) ~ Sexo +  
Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia, data = data_CCR,  
dist = "weibull")
```

Coefficients:

(Intercept)	SexoMujer
-1.8336601	0.4610655
Edad.en.fecha.de.cirugia.>70 años	Edad.en.fecha.de.cirugia.50-60 años
-0.7410115	-0.8072445
Edad.en.fecha.de.cirugia.60-70 años	Quimioterapia
-1.2992278	1.3484816
Radioterapia	
1.6916414	

Scale= 0.612958

Loglik(model)= -49.1 Loglik(intercept only)= -68.8

Chisq= 39.43 on 6 degrees of freedom, p= 5.9e-07

n=82 (6 observations deleted due to missingness)

Call:

```
survreg(formula = Surv(Tiempo_Rec, Evento_1) ~ Sexo +  
Edad.en.fecha.de.cirugia. +  
Quimioterapia + Radioterapia, data = data_CCR, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-1.834	0.680	-2.70	0.007018
SexoMujer	0.461	0.342	1.35	0.177282
Edad.en.fecha.de.cirugia.>70 años	-0.741	0.599	-1.24	0.216061
Edad.en.fecha.de.cirugia.50-60 años	-0.807	0.533	-1.51	0.129849
Edad.en.fecha.de.cirugia.60-70 años	-1.299	0.501	-2.59	0.009571
Quimioterapia	1.348	0.432	3.12	0.001779
Radioterapia	1.692	0.439	3.85	0.000117
Log(scale)	-0.489	0.209	-2.35	0.018983

Scale= 0.613

Weibull distribution

Loglik(model)= -49.1 Loglik(intercept only)= -68.8

Chisq= 39.43 on 6 degrees of freedom, p=  
5.9e-07

Number of Newton-Raphson Iterations: 7

n=82 (6 observations deleted due to missingness)

Se evalúa el modelo de Weibull para las dos situaciones, cuando el parámetro forma  $p > 1$ , el riesgo aumenta a medida que aumenta el tiempo transcurrido, si  $p = 1$  (modelo exponencial) el riesgo es constante y si  $p < 1$  el riesgo disminuye con el tiempo. Otro dato importante para analizar el modelo es la significación del parámetro “scale”, este es el inverso de  $p$ .

Para la situación 1, el output es  $(1 / 0.497 = 2.01)$ , este valor indica que bajo las condiciones de este modelo, el riesgo aumenta a medida que aumenta el tiempo de seguimiento. Observamos que la estimación del parámetro de forma es significativo ( $p = 5.47e-07$ ). Todas las variables tienen significación estadística sobre el modelo. Observamos que el tiempo medio de supervivencia para las mujeres es 0.514 veces el tiempo medio de supervivencia de los hombres.

Para la situación 2, el output es igual  $(1 / 0.613 = 1.631)$ . Este valor indica que bajo las condiciones de este modelo, el riesgo de recaída aumenta a medida que aumenta el tiempo de seguimiento. La estimación del parámetro es significativo ( $p = 0.018983$ ). El tiempo medio de recaída para las mujeres es 0.461 veces el tiempo medio de recaída de los hombres.

### **Comparación entre modelos paramétricos**

Por último evaluamos un modelo que se ajusta mejor a las dos situaciones. La evaluación se lleva a cabo mediante el criterio estadístico de información de “AIC” AKIAKE (Pourhoseingholi, M, Hajizadeh, E., et al 2007) (Bradburn, M, Clark T., et al 2003)

**Tabla 5: Situación 1**

Tiempo/ evento	Modelo	AIC
Situación 1	Exponencial	250.6695
Situación 1	Weibull	193.1406

**TABLA 6: Situación 2**

Tiempo/ evento	Modelo	AIC
Situación 2	Exponencial	116.5954
Situación 2	Weibull	114.1216

Para la comparación de los diferentes modelos nos centramos en los valores obtenidos de AIC. Un valor menor en el criterio AIC representan mejor rendimiento en el modelo. Entre la comparación del modelo Exponencial y Weibull podemos observar que para ambas situaciones el menor valor es para el modelo de Weibull.

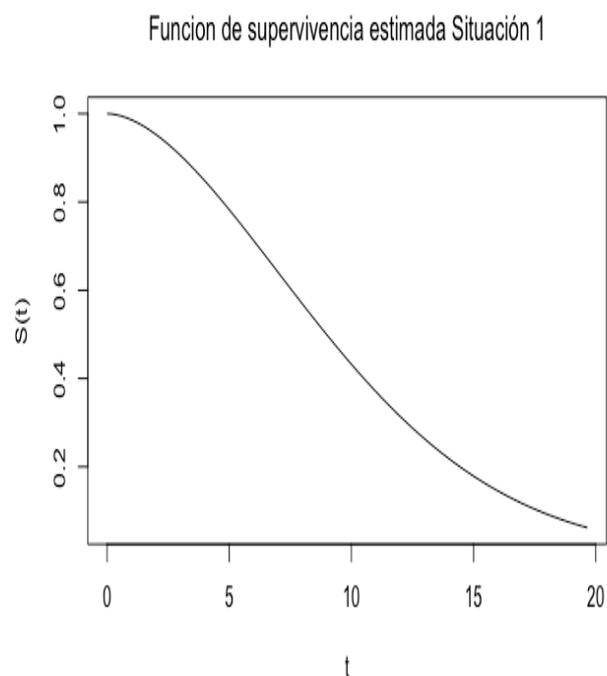
## Comparación entre modelos paramétricos y modelo de Cox

En este apartado se realizan las comparaciones del modelo paramétrico escogido en el epígrafe anterior y el modelo de Cox para las dos situaciones. Teóricamente escoger un modelo de análisis de supervivencia más adecuado es un problema no resuelto. Dependiendo del propósito del estudio puede ser más adecuado un modelo paramétrico o un modelo de Cox.

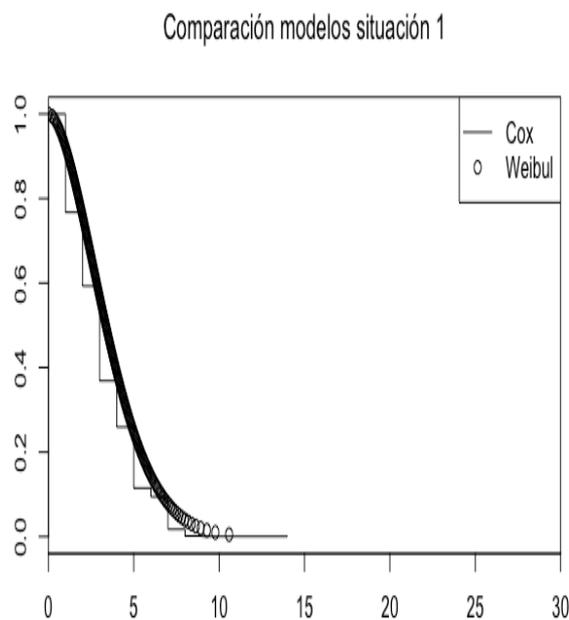
Si se desea comparar riesgos entre distintos niveles de covariables, la teoría aconseja el uso de modelo de Cox, esta situación es la más habitual en estudios médicos. Pero si el interés de la investigación esta basada en obtener información asociada a los parámetros estadísticos como medias, variantes entre otras, como suele ser el interés del área de la industria, probablemente se recomienda el uso de modelo paramétricos (Pourhoseingholi, M, Hajizadeh, E., et al 2007) Swaminathan, R & Brenner, H. 2011)

En el programa R, resultó ser difícil encontrar algoritmos o paquetes que puedan comparar estas situaciones entre modelos. Sin embargo, una opción teórica (Kleinbaum, D & Klien, M) propone comparar las funciones de supervivencia ajustadas dentro de un mismo gráfico. Se realizó la evaluación de comparación con la situación 1.

Se evalúa la función de supervivencia obtenida con el modelo Weibull.



**Figura 22:** Función de supervivencia Weibull



**Figura 23:** Comparación modelos Cox y Weibull

---

La Figura 22 representa la función de supervivencia estimada para la situación 1 con el modelo de Weibull. La Figura 23 representa la comparación entre modelos de Cox y Weibull. Observamos que hay una cierta similitud en las predicciones entre los modelos.

## 2.5 Análisis de resultados

Mediante las diferentes técnicas de análisis de supervivencia implementado el programa R, este trabajo presenta la estimación de las funciones de supervivencia sobre las variables categóricas para la base de datos de pacientes con CCR. La evaluación de covariables predictoras de la función de supervivencia y la verificación de los supuestos del modelo de Cox. Así mismo una evaluación de los modelos paramétricos.

Se realizó un análisis de las variables asociada al tiempo. Este análisis se llevo a cabo mediante la segmentación por situaciones y después un desarrollo del análisis estadístico sobre las mismas. En la primera situación, se evaluaron las funciones de supervivencia tomando como variable (respuesta) el tiempo de supervivencia y como variable censura se utilizó el evento de exitus por enfermedad. Para la segunda situación, se tomó como variable (respuesta) el tiempo hasta la recidiva o recaída del paciente en la enfermedad. Inicialmente se evaluaron las funciones de supervivencia globales para las dos situaciones. Se obtuvo que la probabilidad de supervivencia es bastante escalonada en los primeros 5 años, después se presenta un leve descenso en los eventos, pero luego se torna de nuevo la tendencia disminuida escalonada. La situación 2 sin embargo, presenta una probabilidad de recaída que no alcanza a llegar al 50% en el mismo de tiempo transcurrido del estudio.

Se evaluó en este apartado, la influencia de la variable género en compararon con las demás variables explicativas. El resultado es fiel a la bibliografía con estudios similares, donde sitúa al grupo de los hombres con mayor vulnerabilidad a padecer este tipo de cáncer y su efecto es más complejo con respecto a las variables evaluadas. En ambas situaciones las mujeres tienen mejor pronóstico de supervivencia y recaída.

Una vez ajustado el modelo, la interpretación del modelo de Cox obtenido se realiza mediante el exponencial del coeficiente estimado  $\exp(\beta)$ . Para las variables dicotómicas, es una estimación de la razón de riesgo (hazar ratio) y se interpreta como la cantidad de riesgo obtenida para cada covariable en relación con la ausencia del resto de las covariables. Para el caso de las covariables continuas  $\exp(\beta)$  representa la razón de riesgo al incrementar en una unidad la

---

covariable. Las variables dicotómicas que resultaron ser más determinantes fueron radioterapia y quimioterapia. Para el resto de variables y conforme a la bibliografía se realizó una transformación por rangos en la edad y una selección de los tumores más recurrentes en (tipo de tumor). Se evaluaron y ajustaron en un modelo más específico seleccionando los rangos de edad y los tumores cuyos p-valor fueran más significativos. Los modelos resultantes cumplen con los objetivos del proyecto, se creó un modelo para la base de datos que puede explicar estadísticamente la influencia de las variables sobre el estudio.

Para la comparación de modelos paramétricos, nos encontramos ante la situación de los valores perdidos de la base de datos, valores donde el tiempo es igual a cero. Estos valores no son compatibles cuando se pretende asociar los datos a una distribución Exponencial o Weibull. Para estos resultados tendremos que tomar medidas de eliminación de individuos donde los tiempos son iguales a cero, lo cual desestructura la base de datos original. La teoría sugiere que la identificación del modelo paramétrico más adecuado resulta ser un asunto que requiere de cierto entrenamiento, donde se utiliza una comparación visual de la función de riesgo empírica con la función de riesgo teórica (pueden ser funciones de distribución o densidad). Una opción viable para la comparación de modelos utilizada en la bibliografía es mediante el criterio estadístico AIC, donde obtuvimos que el modelo que mejor se ajustaba a la base de datos es el Weibull.

El programa estadístico R, demostró ser una gran ayuda para el desarrollo del análisis de supervivencia, cuenta con una cantidad considerable de bibliografía e investigaciones que dan soporte a este trabajo. Además, cuenta con la actualización constante de los paquetes utilizados y respaldo en foros y blogs donde se pueden resolver dudas tanto en los fundamentos teóricos como prácticos de los algoritmos, los cuales pueden ser adaptados a las necesidades de nuestra base de datos.

---

### 3. Conclusiones

- ✓ Se concluye que la situación de exitus por enfermedad, resulta ser la menos sesgada para evaluar este estudio.
- ✓ Al evaluar la situación de recaída los datos sobre la media e intervalos de confianza devuelven un resultado indeterminado. Esto se debe a que el ultimo tiempo observado corresponde a una observación censurada, por ende, la estimación de la función de supervivencia no llega a cero. Su valor más pequeño es el estimado en el ultimo tiempo de supervivencia observado. en este caso la estimación se considera infinita más allá del último tiempo que se observó.
- ✓ En general para la dos situaciones, el grupo de mujeres presenta mejor estadística de supervivencia en comparación con el grupo de hombres para esta patología.
- ✓ Para el ajuste de modelo de Cox, es necesario realizar una transformación sobre las variables explicativas respaldada mediante documentos bibliográficos. Es el caso de nuestra base de datos para la variable edad, la cual fue enfocada en pacientes con rango de edad más vulnerables.
- ✓ Para la variable tipo de tumor en la primera situación, el tipo de tumor más significativo resultó ser el no clasificado.
- ✓ En la situación de recaída se pudo realizar un buen ajuste de un modelo significativo. Sin embargo, ocurren casos que tras comprobar las suposiciones de riesgos proporcionales del modelo de Cox, éste no cumple con dichas suposiciones. En la situación 1, nos encontramos que la variable quimioterapia no cumplía con esta suposición, por tanto, el modelo obtenido no se ajustaba a los datos del análisis de supervivencia. La vía de solución a este problema fue utilizar un modelo estratificado de Cox en donde se estratifico la variable quimioterapia. El modelo resultante, cumplía la hipótesis de riesgos proporcionales.
- ✓ Los métodos no paramétricos resultaron ser los más eficientes para realizar el análisis de supervivencia sobre este estudio. La evaluación de métodos paramétricos conlleva a la transformación de los datos, lo cual puede ocasionar estadísticas sesgadas sobre el estudio.
- ✓ Por último, se pudo comprobar que el programa R y sus diferentes paquetes, son un aporte de gran valor para el análisis estadístico de eventos asociados a variables de tiempo.

---

## 4. Glosario

- **IARC**: *International Agency for Research on Cancer*
- **CCR**: *Carcinoma de Células Renales*
- **CCC**: *Carcinoma Renal de Células Claras*
- **CRP**: *Carcinoma Renal Papilar*
- **CRI**: *Carcinoma de células renales inclasificable*
- **CCR**: *Carcinoma cromóforo*

---

## 5. Bibliografía

- ❖ A Torrelles, M Garc, u n Armada, JR Beltr, Moreno, J. V., u n de Laorden, C Sanju, & Rentero, D. B. (2006). An u lisis de supervivencia de los carcinomas de c   lulas renales mrgano-confinados. valor pronmstico de la clasificaci n TNM 1997. *Actas Urolmgicas Espa olas*, 30(7), 655-660.
- ❖ Armesto, D., & Espa a, B. (2011). An lisis de supervivencia. *Electronic Journal of Biomedicine*, (2)
- ❖ Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: Multivariate data analysis--an introduction to concepts and methods. *British Journal of Cancer*, 89(3), 431-436. doi:10.1038/sj.bjc.6601119 [doi]
- ❖ Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part III: Multivariate data analysis -- choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, 89(4), 605-611. doi:10.1038/sj.bjc.6601120 [doi]
- ❖ Brenner, H., & Hakulinen, T. (2002). Up-to-date long-term survival curves of patients with cancer by period analysis. *Journal of Clinical Oncology*, 20(3), 826-832.
- ❖ Borges, R. (2017). Analisis de supervivencia basico utilizando el lenguaje R. [online] Academia.edu. Available at: [http://www.academia.edu/3089513/Analisis\\_de\\_supervivencia\\_basico\\_utilizando\\_el\\_lenguaje\\_R](http://www.academia.edu/3089513/Analisis_de_supervivencia_basico_utilizando_el_lenguaje_R) [Accessed 10 Dec. 2017].
- ❖ Blog.nus.edu.sg. (2017). ST3242.2010/1.2: Introduction to survival analysis | alex.cook. [online] Available at: <http://blog.nus.edu.sg/alexcook/teaching/st3242-20101-2/> [Accessed 15 Nov. 2017].

- 
- ❖ Cancer.org. (2017). American Cancer Society | Information and Resources about for Cancer: Breast, Colon, Lung, Prostate, Skin. [online] Available at: <https://www.cancer.org/> [Accessed 5 Dec. 2017].
  - ❖ Cancer.Net. (2017). Cáncer de riñón: Factores de riesgo y prevención | Cancer.Net. [online] Available at: <https://www.cancer.net/es/tipos-de-cancer/cancer-de-riñon/factores-de-riesgo-y-prevencion> [Accessed 4 Dec. 2017].
  - ❖ Cáncer de Riñón España. (2017). 1.2. Tipos de cáncer renal - Cáncer de Riñón España. [online] Available at: <http://cancer-renal.org/informacion-medica/1-2-tipos-de-cancer-renal/> [Accessed 17 Nov. 2017].
  - ❖ Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232-238. doi:10.1038/sj.bjc.6601118 [doi]
  - ❖ Dalgaard, P. (2008). *Introductory statistics with R* Springer Science & Business Media.
  - ❖ Dawson-Saunders, B., & Trapp, R. G. (1997). *Bioestadística médica* Manual moderno México^ eDF DF.
  - ❖ De Angelis, R., Capocaccia, R., Hakulinen, T., Soderman, B., & Verdecchia, A. (1999). Mixture models for cancer survival analysis: Application to population-based data with covariates. *Statistics in Medicine*, *18*(4), 441-454.
  - ❖ de Vicente, J. G., González, L. L., Arrontes, D. S., Robles, V. C., Cagigal, I. R., González, E. R., & Sánchez, A. B. (2003). Análisis epidemiológico del carcinoma renal de células claras y cromofílico. *Actas Urológicas Españolas*, *27*(1), 22-25.
  - ❖ Fernández, P. (1995). Análisis de supervivencia. *Unidad De Epidemiología Clínica*,
  - ❖ Fizazi, K., Scher, H. I., Molina, A., Logothetis, C. J., Chi, K. N., Jones, R. J., . . . Saad, F. (2012). Abiraterone acetate for treatment of metastatic castration-resistant prostate

- 
- cancer: Final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *The Lancet Oncology*, 13(10), 983-992.
- ❖ Flores-Luna, L., Salazar-Martínez, E., Duarte-Torres, R. M., Torres-Mejía, G., Alonso-Ruiz, P., & Lazcano-Ponce, E. (2008). Factores pronósticos relacionados con la supervivencia del cáncer de mama. *Salud Pública De México*, 50(2), 119-125.
  - ❖ Gomez Perez, L., Budia Alba, A., Delgado Oliva, F. J., Ruiz Cerda, J. L., Trassiera Villa, M., & Jimenez Cruz, F. (2007). Incidental renal cancer in young adults: Clinical and histopathological findings. [Cancer renal incidental en pacientes menores de 40 anos: hallazgos clinicos e histopatologicos] *Actas Urologicas Espanolas*, 31(3), 244-249. doi:S0210-4806(07)73629-X [pii]
  - ❖ Gómez, G., & Langohr, K. (2004). Análisis de supervivencia. *Barcelona, España*,
  - ❖ Guillermo, G. S., Ernesto, J., & Jaime, G. (2010). Carcinoma renal en menores de 40 años; características clínicas e histopatológicas. *Rev Mex Urol*, 70(5), 283-287.
  - ❖ Kleinbaum, D. G., & Klein, M. (2010). *Survival analysis* Springer.
  - ❖ Lambert, P. C., Thompson, J. R., Weston, C. L., & Dickman, P. W. (2006). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3), 576-594.
  - ❖ Le, C. T. (1997). *Applied survival analysis* Wiley.
  - ❖ López-Jiménez, F., Martínez Moreno, M., Obrador Vera, G. and Martínez Gonzales, M. (2005). Manual de medicina basada en la evidencia. México: Editorial el Manual Moderno, p.Capitulo 7 - Análisis de supervivencia y análisis multivariado.
  - ❖ Ljungberg, B., Cowan, N., Hanbury, D., Hora, M., Kuczyk, M., Merseburger, A., . . . Sinescu, I. (2010). Guía clínica sobre el carcinoma renal. *European Association of Urology*, 15

- 
- ❖ Martínez González, M. and de Irala–Estévez, J. (2017). Análisis de supervivencia y análisis multivariado. 17th ed. [ebook] Editorial El manual moderno, pp.1-21. Available at: <https://scholar.google.es> = [Accessed 5 Nov. 2017].
  - ❖ Molina Arias, M. (2015). Hazard ratio: Cuando el riesgo varía a lo largo del tiempo. *Pediatría Atención Primaria*, 17(66), 185-188.
  - ❖ Moore, D. (2016). Applied survival analysis using R. 1st ed. Piscataway,NJ,USA: Springer.
  - ❖ Morselli, E., Frank, A. P., Santos, R. S., Fátima, L. A., Palmer, B. F., & Clegg, D. J. (2016). Sex and gender: Critical variables in pre-clinical and clinical medical research. *Cell Metabolism*, 24(2), 203-209.
  - ❖ Pardo, C., & Cendales, R. (2009). Survival analysis of cervical cancer patients. *Biomédica*, 29(3), 437-447.
  - ❖ Pena, R. E. B. (2005). Análisis de sobrevivencia utilizando el lenguaje R. *Paipa, Boyacá, Colombia*,
  - ❖ Pérez-Hoyos, S. (1997). *Análisis de supervivencia* Institut valencià d'estudis en salut pública.
  - ❖ Pohar, M., & Stare, J. (2006). Relative survival analysis in R. *Computer Methods and Programs in Biomedicine*, 81(3), 272-278.
  - ❖ Pourhoseingholi, M. A., Hajzadeh, E., Moghimi Dehkordi, B., Safaee, A., Abadi, A., & Zali, M. R. (2007). Comparing cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific Journal of Cancer Prevention*, 8(3), 412.
  - ❖ Rao, P. (2000). Applied survival analysis: Regression modeling of time to event data. *Journal of the American Statistical Association*, 95(450), 681-681.
  - ❖ Rebas, P. (2005). Basic concepts in survival analysis. [Conceptos basicos del analisis de supervivencia] *Cirugia Espanola*, 78(4), 222-230. doi:13079636 [pii]

- 
- ❖ Rivas-Ruiz, R., Pérez-Rodríguez, M., Palacios, L., & Talaveraa, J. O. (2014). Del juicio clínico al análisis de supervivencia. *Rev Med Inst Mex Seguro Soc*, 52(3), 308-315.
  - ❖ Rodríguez, J., Ferrer, E., Olmos, A., Codina, S., Borrellas, J., & Piera, L. (2001). Análisis de supervivencia del acceso vascular permanente. *Nefrología*, 21(3)
  - ❖ StataCorp, L. (1985). Stata survival analysis and epidemiological tables reference manual. *StataCorp LP, College Station, Texas*,
  - ❖ Stewart, B., & Wild, C. (2014). World cancer report 2014. 2014. *Lyon CEDEX, France*,
  - ❖ Swaminathan, R., & Brenner, H. (2011). Statistical methods for cancer survival analysis. *IARC Scientific Publications*, (162)(162), 7-13.
  - ❖ Tsui, K., Shvarts, O., Smith, R. B., Figlin, R. A., & BELLDEGRUN, A. (2000). Prognostic indicators for renal cell carcinoma: A multivariate analysis of 643 patients using the revised 1997 TNM staging criteria. *The Journal of Urology*, 163(4), 1090-1095.
  - ❖ Views, R. (2017). Survival Analysis with R. [online] R-bloggers. Available at: <https://www.r-bloggers.com/survival-analysis-with-r-3/> [Accessed 7 Nov. 2017].

---

## 6. Anexos

En este anexo, se detalla el código R utilizado para el estudio estadístico del análisis de supervivencia de este proyecto.

```
# Preparación de los datos
```

```
library(readxl)
```

```
Dataset_01 <- read_excel("~/Desktop/Dataset_03.xlsx", sheet =  
"Data_CCR_R")
```

```
# Data.Frame con los datos a utilizar
```

```
data_CCR <- data.frame(Dataset_01)
```

```
# Codificación de Factores y variables
```

```
data_CCR$Sexo <- as.factor(data_CCR$Sexo)
```

```
data_CCR$Tipo.de.tumor <- as.factor(data_CCR$Tipo.de.tumor)
```

```
data_CCR$Estadio <- as.factor(data_CCR$Estadio)
```

```
# Resumen estadístico inicial
```

```
str(data_CCR)
```

```
summary(data_CCR)
```

```
# Descripción variables categoricas
```

```
library(descr)
```

```
crosstab(data_CCR$Tipo.de.tumor, data_CCR$Sexo, type = c("f", "c"), plot  
= FALSE, prop.r = TRUE)
```

```
descr(data_CCR$Tipo.de.tumor)
```

```
library(sjPlot)
```

---

```

#### Tipo de tumor vs género
sjt.xtab(data_CCR$Tipo.de.tumor, data_CCR$Sexo, show.row.prc=TRUE)

### Edad vs Género
sjt.xtab(data_CCR$Edad.en.fecha.de.cirugia.,
data_CCR$Sexo, show.row.prc=TRUE)

### Gráfica de distribución variable género

library(ggplot2)
library(devtools)
library(easyGgplot2)

ggplot2.histogram(data = data_CCR, xName = "Edad.en.fecha.de.cirugia.",
                  groupName = "Sexo", legendPosition="top",
alpha=0.5, brewerPalette="Paired")

# Recodificación rango de edad para descriptivo
library(car)
attach(data_CCR)
data_CCR$Edad.en.fecha.de.cirugia.[Edad.en.fecha.de.cirugia. <=49] <-
"<50 años"
data_CCR$Edad.en.fecha.de.cirugia.[Edad.en.fecha.de.cirugia. >= 50 &
Edad.en.fecha.de.cirugia. <= 60] <- "50-60 años"
data_CCR$Edad.en.fecha.de.cirugia.[Edad.en.fecha.de.cirugia. >= 61 &
Edad.en.fecha.de.cirugia. <= 70] <- "60-70 años"
data_CCR$Edad.en.fecha.de.cirugia.[Edad.en.fecha.de.cirugia. >= 71] <-
">70 años"
detach(data_CCR)

# Trasformar factor edad

data_CCR$Edad.en.fecha.de.cirugia. <-
as.factor(data_CCR$Edad.en.fecha.de.cirugia.)

```

---

```

# Pasar de días a años dividiendo tiempo por 365

data_CCR$Tiempo_Rec <- round((data_CCR$Tiempo_Rec*12)/365,0)
data_CCR$Tiempo_Sup <- round((data_CCR$Tiempo_Sup*12)/365,0)
### Descriptivo para variable recodificada variable de edad

# Paquetes utilizados para survival analysis
library(sjPlot)
sjt.xtab(data_CCR$Edad.en.fecha.de.cirugia.,
data_CCR$Sexo,show.row.prc=TRUE)
library(survival)
library (survey)
library(SurvRegCensCov)
library(KMsurv)
library(nlme)
library(survminer)

#-----reparar_ggsurvplot-----
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")
devtools::install_github("kassambara/survminer", build_vignettes =
FALSE)
library(survminer)
library(ggpubr)
#-----

### 1) No paramétricos - Estimación Kaplan Meier probabilidad de
supervivencia

## descriptivo supervivencia

table(data_CCR$Evento_2)
svf_01 # descriptivo de la estimación global
svf_02 # estimación global por género
Surv (data_CCR$Tiempo_Sup , data_CCR$Evento_2 ) # resumen con (+)
censura de datos

```

---

```

# A situación 1 (time = Tiempo_Sup, Event_2 = muerte)

# función de supervivencia global

svf_01 <- survfit(Surv(data_CCR$Tiempo_Sup, data_CCR$Evento_2)~1)
summary(svf_01)
plot(svf_01, col = "slategray4", main="Función de supervivencia Kaplan-
Meier", ylab="Probabilidad de supervivencia", xlab = "Tiempo(años)")

# función de supervivencia por género

svf_02 <- survfit(Surv(data_CCR$Tiempo_Sup,
data_CCR$Evento_2)~data_CCR$Sexo)
summary(svf_02)

plot(svf_02,col=c(1,2), main="Funcion de supervivencia por género",
xlab = "Tiempo(años)",ylab = "Probabilidad de supervivencia")
lines(svf_02,conf.int=T,col=c(1,2),lty=c(2,2))
legend(0.3, 0.3, c("Hombre", "Mujer"), col=c(1,2),lty=c(1,1))

library(survminer)
library(ggpubr)

ggsurvplot(svf_02, conf.int=TRUE, pval=TRUE, risk.table=TRUE,
  legend.labs=c("Hombre", "Mujer"), legend.title="Sexo",
  palette=c("dodgerblue2", "orchid2"),
  title="Kaplan-Meier Curve for CCR Survival",
  risk.table.height=.15, data = data_CCR)

table(data_CCR$Evento_1)

Surv (data_CCR$Tiempo_Rec , data_CCR$Evento_1 )

svf_03 <- survfit(Surv(data_CCR$Tiempo_Rec, data_CCR$Evento_1)~1)
summary(svf_03)

```

---

```
plot(svf_03, col = "slategray4", main="Función de supervivencia Kaplan-  
Meier", ylab="Probabilidad de supervivencia", xlab = "Tiempo(años)")
```

```
svf_04 <- survfit(Surv(data_CCR$Tiempo_Rec,  
data_CCR$Evento_1)~data_CCR$Sexo)  
summary(svf_04)
```

```
ggsurvplot(svf_04, conf.int=TRUE, pval=TRUE,  
risk.table=TRUE, legend.labs=c("Hombre", "Mujer"), legend.title="Sexo",  
palette=c("seashell3", "olivedrab3"), title="Kaplan-Meier Curve for CCR  
Survival", risk.table.height=.15, data = data_CCR)
```

```
# A situación 2 (time = Tiempo_Rec, Event_1 = recaída)
```

```
-----
```

```
## descriptivo supervivencia
```

```
Surv (data_CCR$Tiempo_Rec , data_CCR$Evento_1 ) # resumen con (+)  
censura de datos  
svf_03 # descriptivo de la estimación global  
svf_04 # estimación global por género  
table(data_CCR$Evento_1)
```

```
# funcion global supervivencia situación 2
```

```
svf_03 <- survfit(Surv(data_CCR$Tiempo_Rec, data_CCR$Evento_1)~1)  
summary(svf_03)
```

```
plot(svf_03, col = "royalblue1", main="Función de supervivencia Kaplan-  
Meier", ylab="Probabilidad de supervivencia", xlab = "Tiempo(años)")
```

```
# función de supervivencia por género
```

```
svf_04 <- survfit(Surv(data_CCR$Tiempo_Rec,  
data_CCR$Evento_1)~data_CCR$Sexo)
```

---

```

summary(svf_04)

plot(svf_04,col=c(1,2), main="Funcion de supervivencia por género",
xlab = "Tiempo(años)",ylab = "Probabilidad de supervivencia")
lines(svf_04,conf.int=T,col=c(1,2),lty=c(2,2))
legend(0.3, 0.3, c("Hombre", "Mujer"), col=c(1,2),lty=c(1,1))

ggsurvplot(svf_04, conf.int=TRUE, pval=TRUE, risk.table=TRUE,
            legend.labs=c("Hombre", "Mujer"), legend.title="Sexo",
            palette=c("seashell13", "olivedrab3"),
            title="Kaplan-Meier Curve for CCR Survival",
            risk.table.height=.15, data = data_CCR)

ur_prueba <- survfit(Surv(Tiempo_Sup, Evento_2) ~ Sexo + Radioterapia +
Quimioterapia, data = X)
ggsurv <- ggsurvplot(sur_prueba, fun = "event", risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())

ggsurv$plot + theme_bw() + theme (legend.position = "right") +
facet_grid(Sexo~.)

ggsurv
###----- //
-----
-----

sur_prueba_01 <- survfit(Surv(Tiempo_Sup, Evento_2) ~ Tipo.de.tumor,
data = data_CCR)
ggsurv_01 <- ggsurvplot(sur_prueba_01, fun = "event", risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())
ggsurv_01

sur_prueba_02 <- survfit(Surv(Tiempo_Sup, Evento_2) ~
Edad.en.fecha.de.cirugia., data = data_CCR)

```

---

---

```

ggsurv_02 <- ggsurvplot(sur_prueba_02, fun = "event", risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())
ggsurv_02

# univariante Cox

str(data_CCR)

covariates <- c("Sexo", "Edad.en.fecha.de.cirugia.", "Quimioterapia ",
"Radioterapia")

univ_formulas <- sapply(covariates,
                        function(x) as.formula(paste('Surv(Tiempo_Sup,
Evento_2)~', x)))

univ_models <- lapply( univ_formulas, function(x){coxph(x, data = X)})
# Extract data
univ_results <- lapply(univ_models,
                        function(x){
                          x <- summary(x)
                          p.value<-signif(x$wald["pvalue"], digits=2)
                          wald.test<-signif(x$wald["test"], digits=2)
                          beta<-signif(x$coef[1], digits=2);#coeficient
beta
                          HR <-signif(x$coef[2], digits=2);#exp(beta)
                          HR.confint.lower <-
signif(x$conf.int[,"lower .95"], 2)
                          HR.confint.upper <-
signif(x$conf.int[,"upper .95"],2)
                          HR <- paste0(HR, " (",
                                      HR.confint.lower, "-",
HR.confint.upper, ")")
                          res<-c(beta, HR, wald.test, p.value)
                          names(res)<-c("beta", "HR (95% CI for HR)",
"wald.test",

```

---

```

        "p.value")
    return(res)
    #return(exp(cbind(coef(x),confint(x))))
  })
res <- t(as.data.frame(univ_results, check.names = FALSE))
as.data.frame(res)
tfm_surv <- survfit(Surv(Tiempo_Rec, Evento_1) ~ Sexo + Radioterapia +
Quimioterapia, data = data_CCR)
ggsurv_tfm <- ggsurvplot(tfm_surv, pval = T,risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())

ggsurv_tfm$plot + theme_bw() + theme (legend.position = "right") +
facet_grid(Sexo~.)

ggsurv_tfm

# Quitar la variable de sexo
tfm_surv_01 <- survfit(Surv(Tiempo_Rec, Evento_1) ~ Sexo +
Edad.en.fecha.de.cirugia., data = data_CCR)
ggsurv_tfm_01 <- ggsurvplot(tfm_surv_01, pval = T,risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())

ggsurv_tfm_01$plot + theme_bw() + theme (legend.position = "right") +
facet_grid(Sexo~.)

ggsurv_tfm_01

tfm_surv_02 <- survfit(Surv(Tiempo_Rec, Evento_1) ~ Sexo +
Tipo.de.tumor, data = data_CCR)
ggsurv_tfm_02 <- ggsurvplot(tfm_surv_02, fun = "event", pval =
T,risk.table = TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme =
theme_bw())

```

---

```

ggsurv_tfm_02$plot + theme_bw() + theme (legend.position = "right") +
facet_grid(Sexo~.)

ggsurv_tfm_02

# Seleccionar varias observaciones dentro de una columna de variable de
data.frame
tip_tumor <- data.frame(subset(data_CCR, Tipo.de.tumor %in%
c("CCC","CRP")))

tfm_surv_03 <- survfit(Surv(Tiempo_Rec, Evento_1) ~ Tipo.de.tumor,
data = tip_tumor)
ggsurv_tfm_03 <- ggsurvplot(tfm_surv_03, pval = T,risk.table =
TRUE,risk.table.col="strata",conf.int = TRUE, ggtheme = theme_bw())

ggsurv_tfm_03$plot + theme_bw() + theme (legend.position = "right") +
facet_grid(Sexo~.)

ggsurv_tfm_03

summary(tfm_surv_03)
tfm_surv_03

str(data_CCR)

# Modelos

model_1 <- coxph(Surv(Tiempo_Sup,Evento_2)~ Sexo +
Edad.en.fecha.de.cirugia. + Tipo.de.tumor + Quimioterapia +
Radioterapia, data = data_CCR )
model_1

summary(model_1)

```

---

```
model_2 <- coxph(Surv(Tiempo_Sup,Evento_2) ~ strata(Quimioterapia) +  
Radioterapia + Edad.en.fecha.de.cirugia., data = data_CCR)
```

```
model_2  
summary(model_2)
```

```
plot(model_2)  
Comp_model_2 <- cox.zph(model_2)
```

```
Comp_model_2
```

```
ggcoxzph(Comp_model_2)
```

```
ggforest(model_2)
```

```
ggcoxdiagnostics(model_2,type = c("score"))
```

```
modelo_sit_2 <- coxph(Surv(Tiempo_Rec, Evento_1)~  
Sexo+Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia +  
Tipo.de.tumor, data = data_CCR)
```

```
modelo_sit_2  
summary(modelo_sit_2)
```

```
modelo_sit_3 <- coxph(Surv(Tiempo_Rec, Evento_1)~ Sexo+  
(Edad.en.fecha.de.cirugia.=="60-70 años") + Quimioterapia +  
Radioterapia + (Tipo.de.tumor=="CRI"), data = data_CCR)
```

```
modelo_sit_3  
summary(modelo_sit_3)  
cox.zph(modelo_sit_3)
```

```
Comp_model_3 <- cox.zph(modelo_sit_3)
```

---

```

Comp_model_3

ggcoxzph(Comp_model_2)

ggforest(modelo_sit_3)

ggcoxdiagnostics(model_2,type = c("deviance"))

covariates <- c("Sexo", "Quimioterapia", "Radioterapia",
"Edad.en.fecha.de.cirugia.")
univ_formulas <- sapply(covariates,
                        function(x) as.formula(paste('Surv(Tiempo_Rec,
Evento_1)~', x)))

univ_models <- lapply( univ_formulas, function(x){coxph(x, data =
data_CCR)})

# modelo paramétrico
##Eliminar filas con NA##
data_CCR<- data_CCR[!(is.na(data_CCR$Tiempo_Sup)),]

## Eliminar filas con valores = 0
data_CCR<- data_CCR[!(data_CCR$Tiempo_Sup == "0"),]

exp_model_01 <- survreg(Surv(Tiempo_Sup, Evento_2) ~
Sexo+Edad.en.fecha.de.cirugia.+Quimioterapia+Radioterapia , data =
data_CCR, dist = "exponential")

exp_model_01
summary(exp_model_01)

round(exp(summary(exp_model_01)$coef),3)

```

---

```

modelo_00 <-
survreg(Surv(Tiempo_Sup,Evento_2)~Sexo+Edad.en.fecha.de.cirugia.
+Quimioterapia+Radioterapia, data = data_CCR,dist = "weibull")
modelo_00
summary(modelo_00)

## gráfico de la funcion de supervivencia estimada
alfa <- modelo_00$icoef[[1]]
sigma <- modelo_00$scale
c(1/sigma, exp(alfa))
media <- exp(alfa)*gamma(1 + sigma)
S <- function(t) exp(-(exp(-alfa)*t)^(1/sigma))
curve(S, 0, 2*media,
      main="Funcion de supervivencia estimada Situación 1",
      xlab="t", ylab="S(t)")

data_CCR<- data_CCR[!(data_CCR$Tiempo_Rec == "0"),]

exp_model_02 <- survreg(Surv(Tiempo_Rec, Evento_1) ~ Sexo +
Edad.en.fecha.de.cirugia. + Quimioterapia + Radioterapia , data =
data_CCR, dist = "exponentia")
exp_model_02
summary(exp_model_02)
round(exp(summary(exp_model_02)$coef),3)

modelo_01 <-
flexsurvreg(Surv(Tiempo_Rec,Evento_1)~Sexo+Edad.en.fecha.de.cirugia.
+Quimioterapia+Radioterapia, data = data_CCR,dist = "weibull")
modelo_01
summary(modelo_01)

summary(factor(Dataset_01$Evento_2))
summary(factor(Dataset_01$Evento_1))

```

---

```
Dataset_01$Tiempo_Rec <- round(Dataset_01$Tiempo_Rec/365,0)
data_CCR$Tiempo_Sup <- round(data_CCR$Tiempo_Sup/365,0)

Surv(Dataset_01$Tiempo_Rec ,Dataset_01$Evento_1)
Dataset_01$Evento_2

# Comparción entre modelo de cox y weibull, para poner una sola linea,
quitamos la estratificación en la variable quimioterapia

patron1 <- data.frame(Sexo="Mujer", Edad.en.fecha.de.cirugia.="60-70
años",Quimioterapia=1,Radioterapia=1)

plot(survfit(model_2,newdata = patron1),conf.int=FALSE,xlim=c(0,30),
main="Comparación modelos situación 1")
ptc <- seq(0,1,length=200)
days <- predict(modelo_00, newdata = patron1, type = "quantile", p =
ptc)
survival <- 1 - ptc
points(days,survival)
legend("topright",c("Cox","Weibul"),lty=c("solid",NA),pch=c(NA,1))
```