



# **Análisis de supervivencia de pacientes intervenidos quirúrgicamente por hepatocarcinoma**

**Paula Ortega Rioja**  
Máster en Bioinformática y Bioestadística  
Área de Estadística y Bioinformática

Director: **Fernando Carmona Berraquero**

05/2018



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Análisis de supervivencia de pacientes intervenidos quirúrgicamente por hepatocarcinoma</i>
<b>Nombre del autor:</b>	<i>Paula Ortega Rioja</i>
<b>Nombre del consultor/a:</b>	<i>Nombre y dos apellidos</i>
<b>Nombre del PRA:</b>	<i>Fernando Carmona Berraquero</i>
<b>Fecha de entrega (mm/aaaa):</b>	06/2018
<b>Titulación:</b>	<i>Máster en Bioinformática y Bioestadística</i>
<b>Área del Trabajo Final:</b>	<i>Área de Estadística y Bioinformática</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>Análisis de supervivencia, hepatocarcinoma, análisis de datos censurados.</i>

### **Resumen del Trabajo (máximo 250 palabras):**

Los análisis de supervivencia permiten estudiar el periodo de tiempo transcurrido desde un momento predeterminado hasta la ocurrencia de un suceso. En este proyecto, se ha evaluado la supervivencia de pacientes intervenidos de hepatocarcinoma (CHC). Esta dolencia está considerada grave y constituye una de las principales causas de muerte relacionadas con el cáncer.

Teniendo en cuenta que no se dispone de una gran cantidad de datos y que gran parte de ellos son censurados, se han estudiado la supervivencia de los pacientes mediante el método de Kaplan-Mayer, que es especialmente útil para estos casos. Además, permite analizar las variables contrastando distintas funciones para cada uno de sus niveles (GRAMATGES, Anissa, 2002). Los mencionados contrastes se han realizado utilizando la prueba Log-Rank.

Utilizando el método de riesgos proporcionales de Cox, se ha planteado y contrastado un modelo de regresión para la supervivencia de los pacientes, en función de variables explicativas de la base de datos.

Los resultados de este análisis muestran que las variables relacionadas con el grado de desarrollo del CHC son las únicas que proporcionan funciones de supervivencia significativamente distintas.

En cuanto a los grupos incluidos en cada una de las variables restantes, como son etiología, sexo, edad y tratamientos recibidos, no presentan diferencias estadísticas entre sus funciones de supervivencia.

**Abstract (in English, 250 words or less):**

Survival analysis allow studying the period of time elapsed from a predetermined moment to the occurrence of an event. In this project, the survival of patients operated on for hepatocarcinoma (HCC) has been evaluated. This disease is considered really serious and one of the leading causes of death related to cancer.

Considering that the database is short, and a large part of its data is censored, it has been studied by Kaplan-Mayer method, which is especially useful for these cases. Besides this, it allows analyzing the variables contrasting different functions for each of its levels (GRAMATGES, Anissa, 2002). The mentioned contrasts have been made using the Log-Rank test.

Using the Cox proportional hazards method, a regression model for the survival of patients has been proposed and contrasted, based on explanatory variables of the database.

The results of this analysis shows that variables related to the degree of development of the HCC, are the only ones that provide significantly different survival functions.

Regarding the groups included in each of the remaining variables, such as etiology, sex, age and treatments received, they do not show statistical differences among their survival functions.

# Índice

<b>1. Introducción</b>	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	2
1.2.1 Objetivo general .....	2
1.2.2 Objetivos específicos .....	2
1.3 Enfoque y método seguido	2
1.4 Planificación del Trabajo	3
1.5 Breve resumen de productos obtenidos	4
1.6 Breve descripción de los otros capítulos de la memoria	4
<b>2. Materiales y Métodos</b>	5
2.1 Marco teórico	5
2.1.1 Teoría del análisis de supervivencia .....	5
2.1.2 Datos censurados .....	6
2.1.3 Métodos no paramétricos.....	7
2.1.4 Métodos paramétricos .....	9
2.2 Base de datos	11
2.3 Análisis descriptivo	12
2.4 Análisis de supervivencia	17
2.4.1 Función de supervivencia global.....	17
2.4.2 Comparaciones por grupos.....	20
2.4.2.1 Hepatocarcinoma.....	21
2.4.2.2 Sexo .....	22
2.4.2.3 Edad .....	23
2.4.2.4 Tratamientos.....	24
2.4.2.5 Etiología.....	25
2.4.2.6 Recidiva.....	26
2.4.3 Modelo de riesgos proporcionales de Cox.....	27
2.4.3.1 Univariante.....	27
2.4.3.2 Multivariante .....	28
2.4.4 Estimación paramétrica de la función supervivencia .....	30
<b>3. Conclusiones</b>	31
<b>4. Glosario</b>	32
<b>5. Bibliografía</b>	33
<b>6. Anexos</b>	35
6.1 Estadios tumorales según la clasificación de la AJCC	35
6.2 Código de análisis compilado por Rmarkdown	36

## Índice de figuras

Figura 1 Diagrama de Gantt del proyecto	4
Figura 2 Histograma de la variable edad	13
Figura 3 Gráfico de barras de las etiologías causantes de CHC	13
Figura 4 Histograma de la variable longevidad	16
Figura 5 Curva de supervivencia	19
Figura 6 Función de riesgo acumulado global	20
Figura 7 Curvas de supervivencia en función de la diferenciación del CHC	21
Figura 8 Curvas de supervivencia en función del estadio tumoral	22
Figura 9 Curvas de supervivencia en función del sexo	22
Figura 10 Curvas de supervivencia en función de la edad	23
Figura 11 Curvas de supervivencia en función de si se recibió tratamiento previo a la cirugía	24
Figura 12 Curvas de supervivencia en función de si se recibió tratamiento tras a la cirugía	24
Figura 13 Curvas de supervivencia en función del consumo de alcohol	25
Figura 14 Curvas de supervivencia en función de si los pacientes presentaban infección vírica	26
Figura 15 Curvas de supervivencia en función de la recidiva	26
Figura 16 Residuos de los parametros mediante DFBeta	30

## Índice de Tablas

Tabla 1 Variables de la base de datos a estudiar	11
Tabla 2 Estado de diferenciación de los HCC de los pacientes (1=Bien, 2=Moderado, 3 = Poco)	12
Tabla 3 Sexo biológico de los pacientes	12
Tabla 4 Porcentajes de los estadios tumorales de los HCC de los pacientes	14
Tabla 5 Resumen de las frecuencias relativas de las variables: Muerte, Recidiva, PreTreat y PostTreat	15
Tabla 6 Resumen estadístico de la variable longevidad	15
Tabla 7 p-valores de los modelos de Cox para cada variable	27

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

El carcinoma hepatocelular (CHC) es un tumor maligno que se desarrolla en el hígado cuando las células de éste comienzan a crecer de forma desmesurada. El CHC es una patología que podríamos calificar de grave dado que constituye una de las principales causas de muerte relacionadas con el cáncer, confiriéndole una relevancia muy significativa a nivel global.

En el caso concreto de España, la incidencia de esta enfermedad se ha estabilizado debido a que la infección viral y el consumo de alcohol, que son las principales causas de cirrosis, están más controlados. Actualmente, la afectación circunscrita a un período de un año es de entre 10 y 11 personas por cada 100.000 habitantes, lo que sitúa a España en una escala intermedia en comparación con el resto de los países. La incidencia de cáncer hepático es más elevada en el continente asiático y países del África subsahariana, debido a una alta diseminación de la hepatitis vírica (Associació catalana de Malalts d'Hepatitis, 2018, <http://asscat-hepatitis.org/hepatocarcinoma/>).

El factor de riesgo más relevante para el desarrollo de un hepatocarcinoma es la cirrosis. Esta patología se origina cuando, debido a lesiones o enfermedades prolongadas como el alcoholismo o la hepatitis, se sustituye tejido hepático por tejido cicatricial, lo que comporta una pérdida de funcionalidad del mismo.

Las hepatitis víricas crónicas pueden desembocar en cirrosis o directamente producir un CHC, siendo las más comunes a nivel mundial la hepatitis B y la C (PERZ, Joseph F, et al., 2006). Este supuesto queda demostrado por la estrecha relación existente entre las zonas geográficas donde la hepatitis tiene una mayor tasa de incidencia de infección y los casos detectados de carcinoma hepatocelular.

La edad de aparición del carcinoma hepatocelular es variable y depende de la región y de las campañas de vacunación que se realicen en la misma. Se relaciona la aparición en personas de menor edad con las zonas en las que la prevalencia de la hepatitis B es mayor, ya que se puede transmitir durante el embarazo y el parto. Actualmente, se estima que en más del 90% de los países se vacuna del VHB a los recién nacidos, y al 70% de ellos se les administran las 3 dosis recomendadas.

La detección precoz del CHC es clave para poder tratarlo de forma efectiva. Los hígados que padecen hepatitis crónica desarrollan progresivamente una cirrosis, lo que conlleva la aparición de nódulos. Estos nódulos pueden ser inicialmente benignos (regenerativos), en función de su tamaño, o bien CHC. Mediante pruebas radiológicas se puede determinar qué clase de nodulación es y actuar en consecuencia.

Una vez detectada la enfermedad, y en función del estado de desarrollo en el que se encuentre, se determina el tratamiento a aplicar. Si el paciente presenta un tumor pequeño y una cirrosis que permite al órgano ser completamente funcional, lo indicado es una intervención quirúrgica para extraer el tumor. Por el contrario, si el tumor tiene un tamaño mayor, se realizan tratamientos de quimioembolización (procedimiento en el que se bloquea el suministro de sangre al tumor después de administrar medicamentos contra el cáncer a los vasos sanguíneos cercanos al tumor), siendo un tratamiento muy focalizado y menos agresivo que la quimioterapia lo que permite una disminución de los efectos secundarios. En el caso de que el cáncer desarrolle metástasis, tan solo se pueden llevar a cabo cuidados paliativos.

El conjunto de datos utilizado en este trabajo consta de 83 pacientes diagnosticados de cáncer de hígado e intervenidos quirúrgicamente a causa del mismo. Con esta información se pretende construir un modelo de supervivencia y posteriormente, comprobar si existen variables que influyan en mayor o menor medida en las probabilidades de sobrevivir de este grupo de pacientes.

## **1.2 Objetivos del Trabajo**

### **1.2.1 Objetivo general**

- Obtener un modelo de supervivencia que explique la evolución de los pacientes incluidos en la base de datos
- Realizar un análisis diferencial para ver si existen factores con mayor influencia en la supervivencia de los pacientes estudiados

### **1.2.2 Objetivos específicos**

- Realizar una revisión bibliográfica sobre el cáncer de hígado y el análisis de supervivencia
- Realizar un análisis descriptivo de los datos de los pacientes a estudiar
- Construir un modelo de supervivencia a partir de la base de datos
- Evaluar el modelo construido y discutir la incidencia de las diferentes variables incluidas en el modelo de supervivencia

## **1.3 Enfoque y método seguido**

El presente proyecto consta de dos fases diferenciadas:

- La recopilación de material bibliográfico actualizado sobre el cáncer de hígado y modelos de supervivencia.



- El análisis de los datos de pacientes intervenidos de CHC, en una primera instancia de forma descriptiva y posteriormente mediante un modelo de supervivencia en el que se evalúan y se valoran las distintas variables disponibles. El software estadístico que se emplea es Rstudio, junto con los paquetes que se precisen en cada caso y la herramienta RMarkdown, que permite presentar el código elaborado y los resultados obtenidos de él.

#### **1.4 Planificación del Trabajo**

Este proyecto se desglosa en las siguientes tareas:

- Buscar información bibliográfica sobre el marco del trabajo
- Ordenar y clasificar los datos para su posterior utilización
- Realizar un análisis descriptivo de la base de datos
- Evaluar los datos mediante un modelo de supervivencia
- Redactar la memoria del trabajo
- Realizar la presentación para la exposición del proyecto
- Defender el trabajo ante el tribunal evaluador

Mediante el programa GantProject se han plasmado las tareas mencionadas anteriormente de forma cronológica en la *Figura 1*, teniendo en cuenta también los siguientes hitos:

1. Revisión bibliográfica
2. Análisis descriptivo de los datos
3. Realización de un modelo de supervivencia
4. Presentación de la memoria
5. Defensa del trabajo

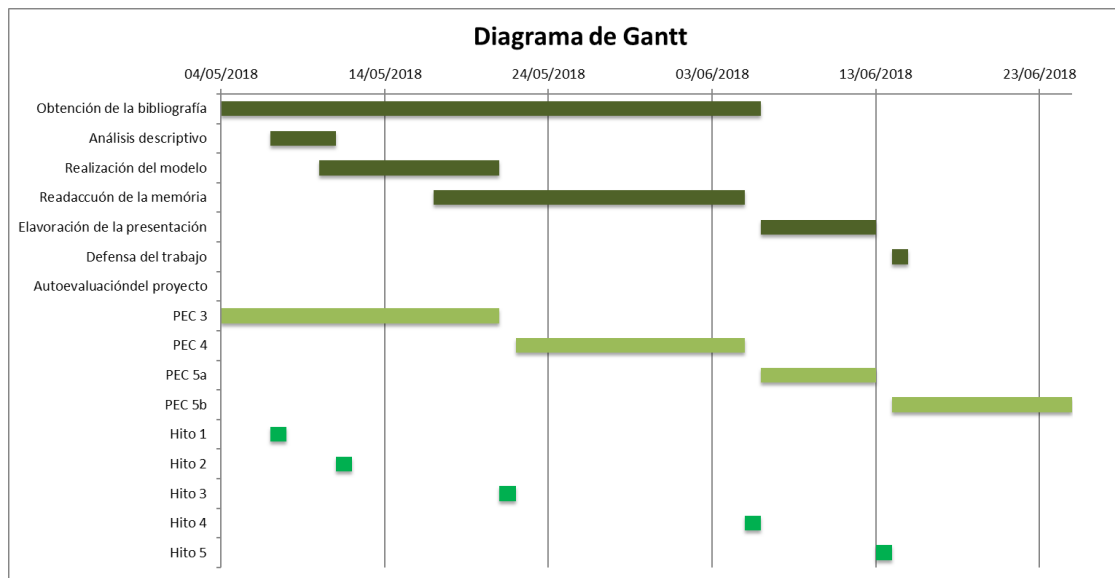


Figura 1 Diagrama de Gantt del proyecto

## 1.5 Breve resumen de productos obtenidos

La realización del presente proyecto desemboca en la obtención de:

Un análisis de supervivencia en función de las variables explicativas de la base de datos y el análisis estadístico mediante métodos no paramétricos, semiparamétricos y paramétricos de los mismos datos. Todo ello reflejado en un archivo Rmarkdown del programa RStudio donde se puede ver tanto el código como los resultados obtenidos.

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

Una vez contextualizado el trabajo, expuestos los objetivos de éste y planificado el desarrollo del mismo, este proyecto consta del apartado de "Materiales y Métodos". En él se expone la teoría del análisis de supervivencia, incluyendo los datos censurados, métodos de análisis paramétricos, métodos no paramétricos y semiparamétricos. Por otra parte, se incluye la descripción y el análisis de la base de datos sobre la que se trabaja en este proyecto.

En última instancia se incluyen las conclusiones extraídas tras el análisis, un glosario que consta de las palabras técnicas utilizadas y la bibliografía recopilada.

## 2. Materiales y Métodos

### 2.1 Marco teórico

#### 2.1.1 Teoría del análisis de supervivencia

El modelo de regresión se utiliza para crear una relación entre una variable resultado y una o más variables independientes. En los modelos de supervivencia, la variable dependiente mide el tiempo hasta la ocurrencia de un suceso de interés y ello nos permite observar cómo afectan las demás variables independientes del modelo al evento a estudiar (HOSMER, David, et al., 2012).

Así pues, la variable respuesta en un análisis de supervivencia tiene dos componentes:

$C_i$  => Es una variable binaria, llamada también estado, que codifica 1 cuando ha ocurrido el evento de interés y 0 cuando no.

$T_i$  => Es el tiempo de seguimiento o supervivencia. Variable aleatoria positiva  $[0, \infty)$ .

La función de supervivencia es la probabilidad de que un individuo sobreviva durante un tiempo superior a  $t$ , expresada como:

$$S(t) = P(T > t)$$

El análisis de supervivencia tiene como principal objetivo la descripción univariante de los datos observados mediante la estimación de la función de supervivencia. Esta función también incluye la velocidad en la que se presenta el evento en el tiempo, es decir la tasa de incidencia de éste (KLEINBAUM, David, et al., 2010).

En los estudios clínicos es de gran utilidad este tipo de análisis, ya que permite estudiar sucesos de interés, en base a la línea temporal, que es la vida de los pacientes. Dicho suceso no queda limitado a términos de vida o muerte, sino a situaciones en la que se mide el tiempo que transcurre desde un momento determinado hasta que sucede un evento de interés. Este, puede ser el tiempo de recurrencia, tiempo que tarda en recibir el alta, tiempo de recuperación tras una intervención, etc. Para un correcto diseño experimental y su posterior análisis, es esencial que el acontecimiento o suceso sujeto a estudio esté perfectamente definido, para poder determinar exactamente la fecha del mismo.

Por todo lo referido anteriormente, resulta de gran importancia monitorear la supervivencia de pacientes con cáncer, creando unos registros de cáncer clínico y poblacional. Para estar lo más actualizado posible, el período incluido en el análisis debe restringirse a un lapso de tiempo relativamente corto. Sin embargo, para los registros de cáncer que cubren poblaciones relativamente pequeñas o para formas raras de

cáncer es correcto ampliar el período a, por ejemplo, 5 años (BRENNER, Hermann, et al., 2006).

Los datos de supervivencia pueden estudiarse también mediante la función de riesgo acumulado. Mientras que la función de supervivencia se centra en el lapso de tiempo de la "no ocurrencia" del evento, la función de riesgo se centra en la "ocurrencia" del mismo. Esta función,  $H(t)$ , es la probabilidad de que a un individuo siendo observado en un tiempo  $t$ , le suceda un evento determinado, dentro de ese período (REBASA, Pere 2005). Cuando esta función es creciente indica que es más probable que el acontecimiento ocurra con el tiempo, en el caso de este proyecto se trata del fallecimiento de los pacientes. La comentada función se expresa bajo la siguiente fórmula:

$$H(t) = \frac{f(t)}{S(t)}$$

Estos análisis permiten también realizar comparaciones entre las diversas variables intervinientes en un estudio, como pueden ser dosificación de medicamentos, posibles recaídas, curas y tratamientos, entre otros. Esta tipología de trabajos es muy útil para poder determinar factores que representen una mejoría para los pacientes o ayudar a predecir evoluciones.

### **2.1.2 Datos censurados**

El seguimiento de los pacientes viene definido por una fecha de inicio y una fecha de cierre que determinan el tiempo de seguimiento. Las fechas de inicio y cierre suelen ser diferentes para cada individuo en ensayos clínicos de seguimiento, pues los pacientes que se incluyen en el estudio no se incorporan en el mismo momento temporal.

El periodo de seguimiento puede terminar por las siguientes razones (FERNÁNDEZ, Pita 1995):

- a. El paciente decide no participar más en el estudio
- b. Se pierde el paciente y por tanto la información de este
- c. El estudio termina antes de aparecer el evento

Cuando se da una pérdida de seguimiento no se conoce el tiempo hasta el suceso de interés por lo que hablamos de datos censurados.

Los datos censurados son, juntamente con los datos truncados, fallas de datos que proporcionan información parcial sobre la variable o variables de interés (GIJBELS, Irène,

2010). Los datos truncados, sin embargo, son aquellos en los que ciertos sujetos son sistemáticamente excluidos de la muestra por alguna causa común.

La pérdida de seguimiento de los pacientes es un ejemplo de dato censurado. En este caso no se puede determinar si el evento de interés se ha dado o no en todo el período de estudio, pero sí se puede afirmar que hasta la fecha que se tiene constancia no se ha producido el suceso.

Los datos pueden ser censurados, bien por la derecha (individuos en los que se ha perdido el seguimiento tras empezar el estudio), bien por la izquierda (los individuos se incorporaron más tarde y falta el control desde que se empezó hasta su incorporación).

Una observación se considera censurada por la derecha cuando el evento de interés no tiene lugar durante el periodo de seguimiento, de tal forma que no es posible determinar el tiempo de ocurrencia de dicho evento. Existen diversas causas para que ello se produzca, por lo que podemos distinguir entre las siguientes clases de censura:

- Censura Tipo I. El final del seguimiento ocurre en un tiempo predeterminado, el cual es prefijado en el diseño experimental
- Censura Tipo II. La finalización del periodo de seguimiento se da en una proporción preestablecida de individuos. En el diseño no contempla el tiempo en que se dejará de realizar el seguimiento, sino que se fija el final del estudio en el momento que un número predeterminado de individuos falle
- Censura Tipo III. Se da una censura aleatoria, durante la realización del estudio y sin fijar previamente ningún parámetro, se pierde el seguimiento por algún motivo (KLEIN, John P, et al., 2005)

Otra posibilidad es que existan datos que presenten una censura por la izquierda. Este caso, es la consecuencia de que un sujeto entre a un estudio una vez transcurrido el suceso de interés. Se sabe que el evento se ha producido, pero no es cuantificable, dado que no se sabe el momento temporal concreto.

Por último, también es posible una censura en un intervalo, que se da en caso de que el suceso ocurra entre dos momentos de tiempo conocidos, pero no sea posible especificar el instante exacto.

### **2.1.3 Métodos no paramétricos**

Los datos censurados son observaciones que poseen información útil, más incompleta, pero no por ello deben ser desestimados (ARRIBALZAGA, Eduardo, 2007). Para poder analizar este tipo de información es necesario un análisis estadístico pertinente. Los estimadores no paramétricos son la opción más utilizada por su sencillez, ya que

permiten aplicarlos directamente sobre las observaciones, en caso de tener una única clase de elemento suceso.

El estimador no paramétrico de Kaplan-Meier es un método que estima las probabilidades de supervivencia  $S(t_i)$  en un periodo de tiempo concreto, dividiendo el tiempo en intervalos y calculando la supervivencia, en cada uno de esos intervalos. Por esta razón, es especialmente útil para estudios que utilizan un número pequeño de pacientes. Este método, utiliza la premisa de una probabilidad condicional compuesta, es decir, para sobrevivir en un momento determinado debe haber sobrevivido en todos los anteriores:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

Donde  $d_i$  es el número de eventos observados en el momento  $t_i$  y  $n_i$  el número de individuos en riesgo en el tiempo  $t_i$ .

Con este método podemos realizar curvas de supervivencia para cada nivel de variable predictora y así comparar el efecto de estas en el desarrollo del suceso que se estudia.

En investigaciones clínicas es frecuente querer comparar la supervivencia de 2 o más grupos de individuos. Los grupos diferirán respecto a cierto factor (edad, sexo, protocolo de tratamiento, etc.) y es el efecto de este factor sobre la supervivencia lo que interesa conocer. Para contrastar si existen diferencias entre los dos grupos, referente a la supervivencia de estos, debemos realizar un test estadístico (GRAMATGES, Anissa, 2002).

El test Log-Rank, compara el número de eventos (muertes, fracasos) de cada grupo, con el número de eventos esperados para el mismo grupo de individuos. Se emplea un estadístico chi-cuadrado para realizar la comparación de los distintos conjuntos que queramos evaluar, tal como se plasma en la siguiente formula:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Donde:

- $O_i$  es el número de eventos observados del grupo  $i$
- $E_i$  es el número de eventos esperados para el grupo  $i$
- $k$  es el número de grupos a comparar
- $\chi^2$  es el estadístico chi-cuadrado con  $k - 1$  grados de libertad

Otra opción posible es el test Wilcoxon, que utiliza ponderaciones según el número de individuos en riesgo. Pese a su existencia, se considera que la prueba de Log-rank es tan robusta, indiferentemente del número de pacientes que se estudie, que la utilización de cualquier otro método debe estar muy bien justificada (REBASA, Pere ,2005).

### 2.1.4 Métodos paramétricos

Los métodos paramétricos para el análisis de modelos de supervivencia tienen como condición necesaria que los datos se ajusten a la distribución de la función a emplear (Gamma, Exponencial, Log-normal, etc.). Las más utilizadas para esta tipología de análisis son:

La función de supervivencia exponencial:  $S(t) = e^{-ct}$

La función de supervivencia Weibull:  $S(t) = e^{-(ct)^\alpha}$

Una distribución exponencial de nuestros datos expresa un riesgo constante y creciente a lo largo del tiempo. La función de Weibull engloba a la exponencial para  $\alpha=1$ , como se puede apreciar en la formula, por lo que es más general.

La principal ventaja de los modelos paramétricos es la especificidad que proporcionan para realizar el modelo, pero para ello se debe conocer la distribución de los datos a utilizar. En general estos modelos se utilizan de forma efectiva cuando el investigador dispone de un conocimiento amplio del proceso sujeto a estudio (HOSMER, David, et al., 2012).

Esta función se puede estimar por Máxima Verosimilitud (MLE) o bien utilizando el estimador Nelson-Aalen  $\tilde{H}(t)$ .

$$MLE = -\ln(S(t)) \tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

Los métodos expuestos anteriormente permiten estimar las funciones de supervivencia para una muestra aleatoria y comparar dos o más estimaciones. El modelo de riesgos proporcionales de Cox, por otro lado, permite plantear un modelo de regresión para la supervivencia en función de variables explicativas, y comparar dichas estimaciones, teniendo en cuenta el efecto de otras variables distintas de la que se utiliza para definir los grupos.

Este método tiene en cuenta además de la producción del suceso, el tiempo que tarda en producirse.

$$\ln(\lambda_t) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Dónde  $\lambda_t$  representa la función de riesgo de que suceda el evento en el momento t.

El modelo plantea el logaritmo del riesgo relativo como una función lineal de las variables independientes. Se supone, por tanto, que el riesgo relativo no depende del tiempo, sino que es constante a lo largo de este.



## 2.2 Base de datos

La base de datos con la que se trabajará en este proyecto consta de una serie de 83 pacientes intervenidos quirúrgicamente y diagnosticados de carcinoma hepatocelular (CHC). En ella se incluyen los datos para cada paciente de las variables que se denotan en la siguiente tabla (*Tabla 1*).

*Tabla 1 Variables de la base de datos a estudiar*

<b>Variable</b>	<b>Niveles</b>	<b>Descripción</b>
HCC	G1: Bien diferenciado G2: Moderadamente diferenciado G3: Poco diferenciado Entre puntos intermedios diferenciación de los mencionados	Diferenciación del Hepatocarcinoma
Etiología	Etanol VHB VHC Combinación de los factores anteriores	Causa del CHC del paciente.
Cirrosis	P: Presencia A: Ausencia	Presencia o no de cirrosis en el hígado del paciente
Estadio	IA II IIA IIIA IVA	Estadio del tumor del paciente, según la escala del <i>American Joint Committee on Cancer (AJCC)</i> , adjunto en los anexos
Cirugía	Hepatectomía Lobectomía Resección Segmentectomía Tumorectomía	Tipo de intervención quirúrgica a la que el paciente fue sometido
Muerte	0: Paciente vivo 1: Paciente fallecido	Exitus o no del paciente durante el seguimiento
Longevidad	Variable continua (Años)	Años de vida desde la operación del paciente
Recidiva	Si No	Reaparición del tumor tras la cirugía
Sexo	Mujer Hombre	Sexo biológico del paciente
Edad	Variable continua (Años)	Edad biológica del paciente en la fecha de la intervención
Pretratamiento	Si No	Tratamientos realizados antes de la cirugía
Postratamiento	Si No	Tratamientos realizados posteriores a la cirugía

## 2.3 Análisis descriptivo

En este apartado se lleva a cabo un resumen descriptivo del conjunto de datos para cada una de las variables descritas en el apartado anterior (2.2).

La detección del hepatocarcinoma es un factor determinante en el diagnóstico del cáncer hepático. Junto con el de páncreas el cáncer de hígado es, a menudo, de los cánceres que más tarde se diagnostica y de los que cuentan con menos tratamientos eficaces. Si éste se detecta en un estado incipiente las posibilidades de curación del paciente, como es lógico, son mucho mayores. En el caso del hepatocarcinoma, tan solo el 10% de los casos se detecta a tiempo para poder aplicar tratamientos curativos (*Associació catalana de Malalts d'Hepatitis, 2018, <http://asscat-hepatitis.org/hepatocarcinoma/>*).

Dado que los datos utilizados para este proyecto, son de pacientes a los que se les practicó una cirugía, entran dentro del diez por ciento de casos diagnosticados de forma suficientemente temprana.

El hepatocarcinoma de estos pacientes se detectó mayoritariamente (prácticamente un 60% de los casos) en un estadio moderadamente diferenciado, tal como se muestra en la *Tabla 2*.

*Tabla 2 Estado de diferenciación de los HCC de los pacientes (1=Bien, 2=Moderado, 3 = Poco)*

HCC	1	1-2	2	2-3	3
Freq.Relativas	17.7%	6%	59.9%	2.5%	16.4%

La incidencia del cáncer hepático en la población, es más frecuente en el sexo masculino que en el femenino, con una tasa de afectación de aproximadamente 3:1 (Ferlay et al., 2010), proporción que a nivel español se mantiene (LópezAbente et al., 2005). En el caso que nos ocupa, los pacientes estudiados presentan una proporción de varones todavía mayor, cerca de 4 varones por cada mujer, como se puede ver en la *Tabla 3*.

*Tabla 3 Sexo biológico de los pacientes*

Sexo	Hombres	Mujeres
Freq.Relativas	80.7%	19.3%

La edad de los pacientes está comprendida entre los 40 y los 80 años, con una media de 58 años. La *Figura 2* muestra la distribución de las frecuencias de esta variable, donde se puede apreciar una franja notoriamente más elevada que va de los 55 a los 65.

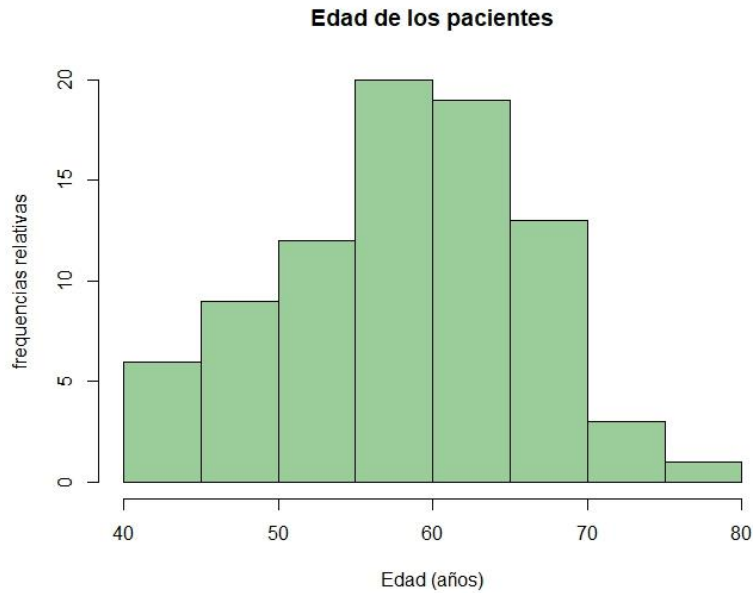


Figura 2 Histograma de la variable edad

Dado que existen causas diversas que pueden producir un cáncer hepático, es relevante conocer las que tienen una mayor incidencia en los pacientes que se evalúan en el proyecto.

La etiología que provoca el carcinoma hepatocelular puede deberse al virus de la hepatitis C o B o al consumo de etanol, o bien, una combinación de ellas, como ya se ha comentado anteriormente. En la *Tabla 4* se muestra la distribución de estas causas en el grupo de pacientes.

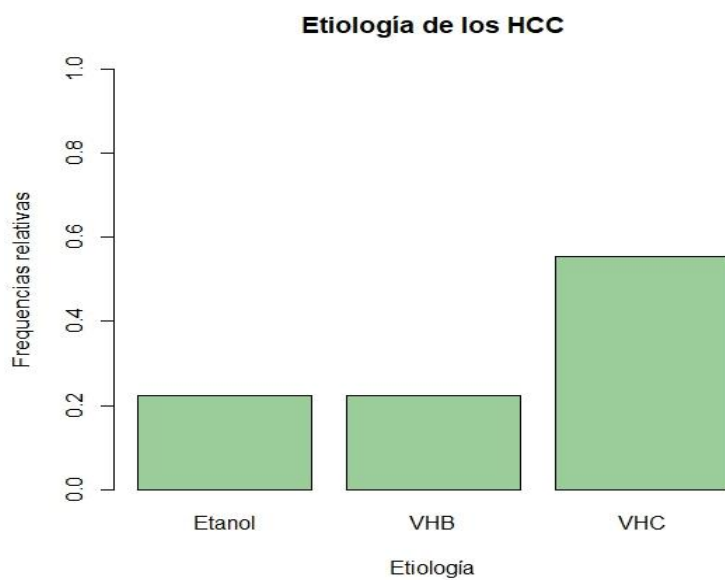


Figura 3 Gráfico de barras de las etiologías causantes de CHC

Como se desprende del gráfico, el virus de la hepatitis C es el que tiene una mayor incidencia (55.5%) como desencadenante de la enfermedad. Con un grado menor de afectación, el virus de la hepatitis B y el consumo de etanol aparecen, en una proporción idéntica (22.2%). Cabe resaltar, que el 90% de los pacientes sujetos a estudio, padecía de cirrosis.

El desarrollo del cáncer hepático se inicia generando nódulos infracentrímetros precancerosos, que finalmente desembocan en tumores malignos. A partir de los tres centímetros, el cáncer ya no puede considerarse inicial y las posibilidades de curación disminuyen.

El estadio tumoral mayoritariamente se encuentra en un estadio II<sup>1</sup> (56.9%), que implica un solo tumor de entre 2 y 5 cm y que no se ha propagado a otras zonas del cuerpo. El 19.4% de los casos presentaban un solo nódulo inferior o igual a 2 cm. El porcentaje se ve levemente incrementado, con un 20.8%, en el caso de los pacientes con varios nódulos (con al menos uno mayor de 5 cm). Estos datos se resumen en la siguiente tabla (*Tabla 4*).

*Tabla 4 Porcentajes de los estadios tumorales de los HCC de los pacientes*

Estadio AJCC	IA	II	IIIA	IIIB	IIVA
Freq.Relativas	19.4%	56.9%	20.8%	1.3%	1.3%

Se puede observar que el grupo de pacientes, oscila en un rango de estadio tumoral entre IA y IIIA, pudiendo presentar tumores más grandes o más pequeños, caracterizados todos ellos por no desarrollar metástasis.

Para que los pacientes puedan ser intervenidos quirúrgicamente, deben cumplir una serie de requisitos que garantizan una tasa de supervivencia superior al 70%, ya que los posibles beneficios derivados de la intervención quirúrgica superan los riesgos inherentes a la misma.

Estos criterios, se basan en que el tumor debe estar formado por un nódulo menor de cinco centímetros o por tres nódulos menores de tres centímetros.

El 65% de las cirugías practicadas en el grupo de estudio, fue una hepatectomía (escisión del hígado) con trasplante ortotópico del hígado (el órgano trasplantado ocupa la posición anatómica original). La mayoría de los pacientes no recibió tratamientos previos a la intervención, pero sí tras ella (*Tabla 5*).

---

1 Descripción más detallada en el anexo

Una vez se ha tratado el cáncer, este puede volver a manifestarse en el mismo órgano o en otras partes del organismo.

En la *Tabla 5*, podemos observar que en la mayoría de los casos los pacientes no presentaron recidiva, durante el tiempo de seguimiento del estudio.

*Tabla 5 Resumen de las frecuencias relativas de las variables: Muerte, Recidiva, PreTreat y PostTreat*

Muerte		Recidiva		PreTreat		PostTreat	
Si	No	Si	No	Si	No	Si	No
33.3%	66.6%	32.5%	67.5%	33.7%	66.2%	59.0%	40.9%

El factor más relevante para el estudio es la variable tiempo, expresada mediante la longevidad de los pacientes. Este intervalo de tiempo abarca desde la intervención quirúrgica hasta el final del estudio o la pérdida del seguimiento.

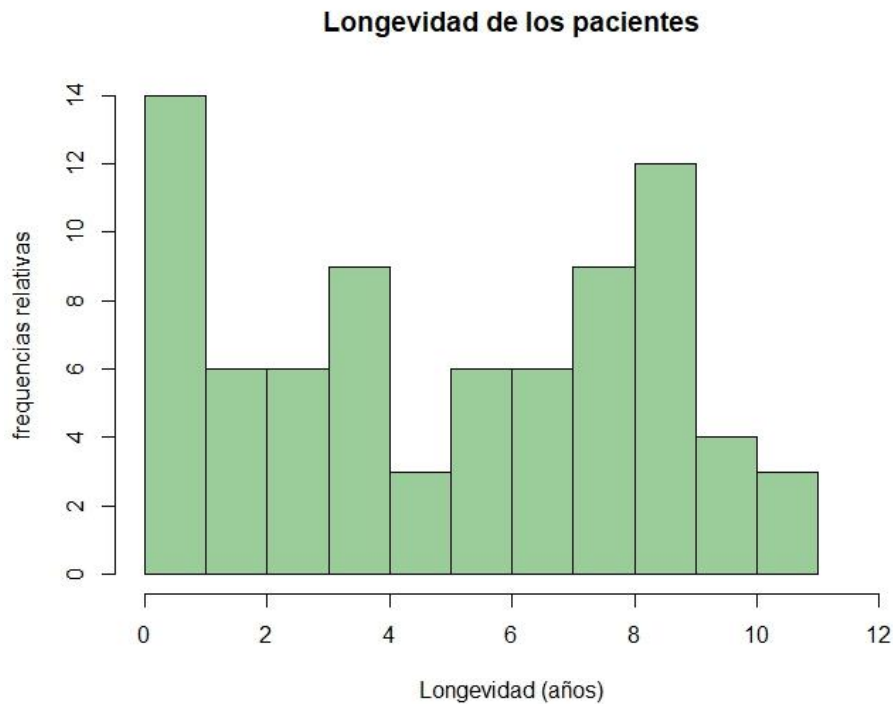
En la *Tabla 6* se resume la línea temporal de los individuos sujetos a estudio. Hay que tener en cuenta que el máximo de 10.5 años expresado no es del todo real, ya que al tener datos censurados puede ocurrir que existan pacientes que hayan tenido una longevidad mayor.

La media de años de vida de los pacientes es de unos 5 años aproximadamente, teniendo en cuenta lo mencionado en el párrafo anterior.

*Tabla 6 Resumen estadístico de la variable longevidad*

Min	1st Qu.	Mediana	Media	3rd Qu.	Max.	NA's
0.08	1.70	5.13	4.96	7.95	10.5	5

La figura (*Figura 4*) siguiente, muestra la distribución de la longevidad de los pacientes, expresada en frecuencias relativas.



*Figura 4 Histograma de la variable longevidad*

La longevidad de los pacientes sujetos a análisis es muy variable. En la *Figura 4*, se observa un pico inicial de frecuencias muy elevado en el que los pacientes no superan el año de vida o bien, se pierde su seguimiento y son datos censurados. Por otra parte, se puede apreciar otro despunte entre los 7 y los 9 años tras la intervención, que podría manifestar un período clave, para comprobar si tras la intervención se ha tenido éxito.

Para los estudios de supervivencia es relevante tener en cuenta los datos en los que se ha perdido el seguimiento, como ya se ha mencionado. En este caso, se trata de datos censurados por la derecha de *Tipo III*, ya que aleatoriamente se ha perdido su seguimiento. En los datos que se analizarán hay 53 valores censurados que suponen un 64% de los datos totales.

En España se ha desarrollado un registro nacional para evaluar las características de los pacientes con CHC. Según estos datos un 89% de los pacientes diagnosticados con esta patología eran cirróticos, el 42% estaban infectados por el virus de la hepatitis C y el 30% consumía alcohol (VARELA, et al., 2010). Al contrastarlo con los datos que se estudian en este proyecto, vemos que la proporción de pacientes cirróticos se confirma con los datos recogidos a nivel nacional. Por otro lado, los porcentajes de infección del VHC y de consumo de etanol también se mantienen similares al registro nacional. Con ello podemos corroborar que la base de datos que se utiliza es una muestra representativa de los pacientes que padecen de cáncer hepático en España. Este hecho, permitirá obtener resultados más sólidos y extrapolables a otros individuos.

## 2.4 Análisis de supervivencia

En primer lugar, se introducen los datos contenidos en un archivo “.xlsx”, en el software R, mediante la función `read_excel()` de la librería `readxl`.

Para poder trabajar con la base de datos, se recodifican las variables y se introduce la tipología de variable que representa cada una, obteniendo como resultado:

```
Classes 'tbl_df', 'tbl' and 'data.frame': 83 obs. of 14 variables:
 $HCC: Factor w/ 5 levels "1","1 y 2","2",...: 1 NA 4 3 1 3 3 3 NA 5 ...
 $Etiologia: Factor w/ 2 levels "NO","SI": NA NA NA NA NA NA NA NA NA N
 A ...
 $Etanol: Factor w/ 2 levels "NO","SI": 1 NA 1 1 1 2 NA 2 1 1 ...
 $Virologia: Factor w/ 2 levels "NO","SI": 2 NA 2 2 2 1 NA 2 2 2 ...
 $Cirrosis: Factor w/ 2 levels "NO","SI": 2 NA 2 2 2 2 2 2 2 2 ...
 $EstadioT: Factor w/ 5 levels "IA","II","IIIA",...: 2 NA 2 3 NA 3 4 2 N
 A 3 ...
 $Cirugia: Factor w/ 25 levels "Hemipatectomía izquierda",...: 7 2 7 7 2
 4 7 7 7 12 16 ...
 $Muerte: num 0 NA 1 0 0 0 0 0 0 0 ...
 $Longevidad: num 3.12 NA 3.45 10.36 8.79 ...
 $Recidiva: Factor w/ 2 levels "NO","SI": 1 NA 2 1 1 1 1 1 2 2 ...
 $PostTreat: Factor w/ 2 levels "NO","SI": 1 1 2 2 1 2 1 2 1 1 ...
 $PreTreat: Factor w/ 2 levels "NO","SI": 1 1 1 1 1 1 1 1 1 1 ...
 $Edad: num 46 55 43 48 54 61 65 53 68 49 ...
 $Sexo: Factor w/ 2 levels "Hombre","Mujer": 1 1 1 1 1 1 1 1 1 1 ...
```

### 2.4.1 Función de supervivencia global

A continuación, se prosigue realizando la función supervivencia de forma global para observar la longevidad de los pacientes. Debido a la presencia de datos censurados, se realiza mediante el método no paramétrico de Kaplan-Meier, con el paquete `survival`.

Previo a la creación de la función, es necesario construir el objeto `Surv`, que combina la información proporcionada por las variables “Longevidad” y “Muerte”. Estas variables suministran la información de, tiempo transcurrido hasta la observación y un indicador dónde 1 es la aparición del evento (el fallecimiento en este caso) y 0 es un dato censurado.

```
[1] 3.12328767+ NA? 3.45205479 10.35616438+ 8.79452055+
 [6] 8.87671233+ 1.47945205+ 10.52054795+ 3.53424658+ 1.47945205+
 [11] 10.10958904+ 9.86301370+ 5.58904110 5.91780822+ 4.93150685
 [16] 1.06849315+ 9.86301370+ 7.97260274+ NA 9.20547945+
 [21] 8.87671233+ 1.31506849+ NA 9.04109589+ 2.95890411
 [26] 8.71232877+ 0.24657534 3.20547945+ 8.71232877+ 0.90410959
 [31] 8.38356164+ 0.16438356 2.95890411 6.49315068+ 2.13698630+
 [36] 8.38356164+ 8.13698630+ 8.21917808+ 7.89041096+ 8.30136986+
 [41] 8.05479452+ 7.47945205+ 0.57534247 0.98630137+ 7.39726027
 [46] 3.69863014+ 4.68493151 8.79452055+ 3.45205479+ 7.72602740
 [51] 3.86301370+ 0.98630137 2.87671233 5.67123288+ 7.23287671+
```

```
[56] 0.41095890+ 6.73972603+ 1.31506849 5.09589041+ 2.79452055
[61] 7.06849315+ 5.17808219+ 7.06849315+ 7.31506849+ 6.82191781+
[66] 6.65753425+ 0.08219178+ 0.65753425+ 6.82191781+ NA
[71] 4.35616438 3.04109589 0.41095890 6.08219178+ 0.32876712
[76] 0.49315068+ 1.56164384 NA? 0.08219178+ 3.28767123
[81] 5.34246575+ 0.73972603 2.63013699
```

Los datos censurados se muestran con un símbolo + que los acompaña a la derecha.

La creación de la función de supervivencia se ejecuta a través de la orden `survfit()` y se visualiza dicha función utilizando `print()`:

```
5 observations deleted due to missingness
  n events median 0.95LCL 0.95UCL
78.00 24.00 NA 7.73 NA
```

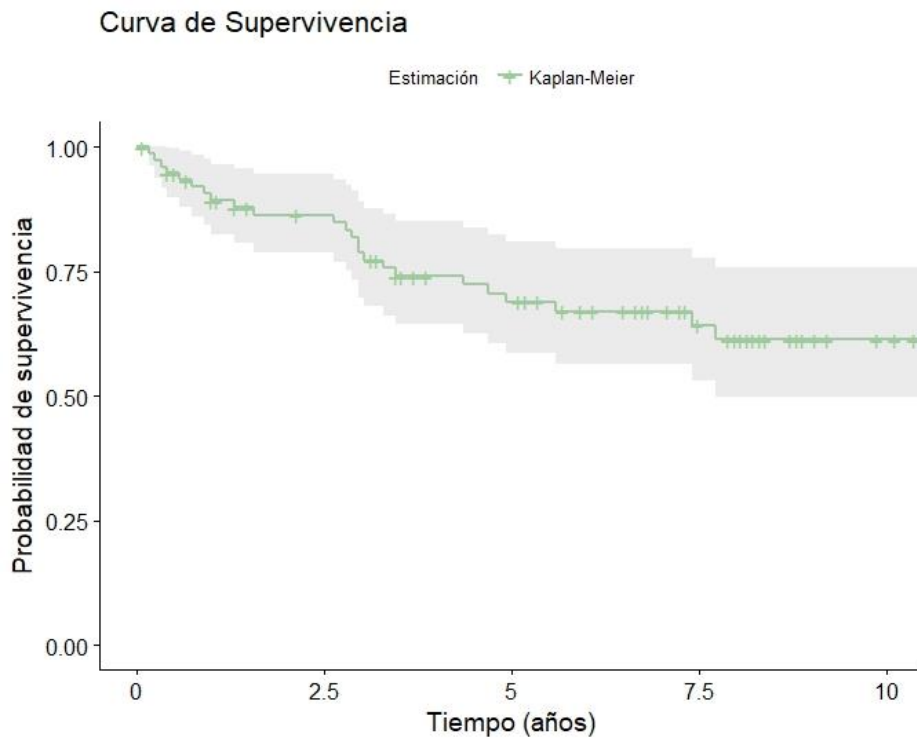
En esta, se puede ver que el número total de pacientes que se consideran para la función es de 78, de los cuales fallecieron 24 durante este seguimiento.

Con `summary()` se observa un resumen de la estimación de los datos de la función, y con ellos se crea la curva, que se muestra en la *Figura 5*, utilizando `ggsurvplot()` que incluye los intervalos de confianza para  $\alpha=0.05$ .

```
5 observations deleted due to missingness
time n.risk n.event survival std.err lower 95% CI upper 95% CI
0.164 76 1 0.987 0.0131 0.962 1.000
0.247 75 1 0.974 0.0184 0.938 1.000
0.329 74 1 0.961 0.0223 0.918 1.000
0.411 73 1 0.947 0.0256 0.898 0.999
0.575 70 1 0.934 0.0286 0.879 0.992
0.740 68 1 0.920 0.0313 0.861 0.984
0.904 67 1 0.906 0.0337 0.843 0.975
0.986 66 1 0.893 0.0359 0.825 0.966
1.315 63 1 0.878 0.0380 0.807 0.956
1.562 59 1 0.864 0.0402 0.788 0.946
```

(...)





*Figura 5 Curva de supervivencia*

La curva de supervivencia no presenta una inclinación muy marcada, sino que desciende paulatinamente hasta el 61% de probabilidad, en el transcurso de los prácticamente 11 años que abarca la gráfica. Los pacientes intervenidos de hepatocarcinoma, según estos datos, tienen una tasa de supervivencia elevada. Según la *Associació Catalana de Malalts d'Hepatitis (ASS)*, entre el 50% y el 70% de los pacientes intervenidos, conseguirán sobrevivir los cinco primeros años. Estos datos corroboran el resultado de la curva obtenida para los pacientes que se estudian en el presente proyecto.

La función de riesgo acumulado, que se muestra en la *Figura 6*, permite ver que el riesgo de fallecer aumenta progresivamente tras la intervención quirúrgica. También se puede apreciar, que entre los dos años y medio y los tres se produce un crecimiento brusco de este riesgo.

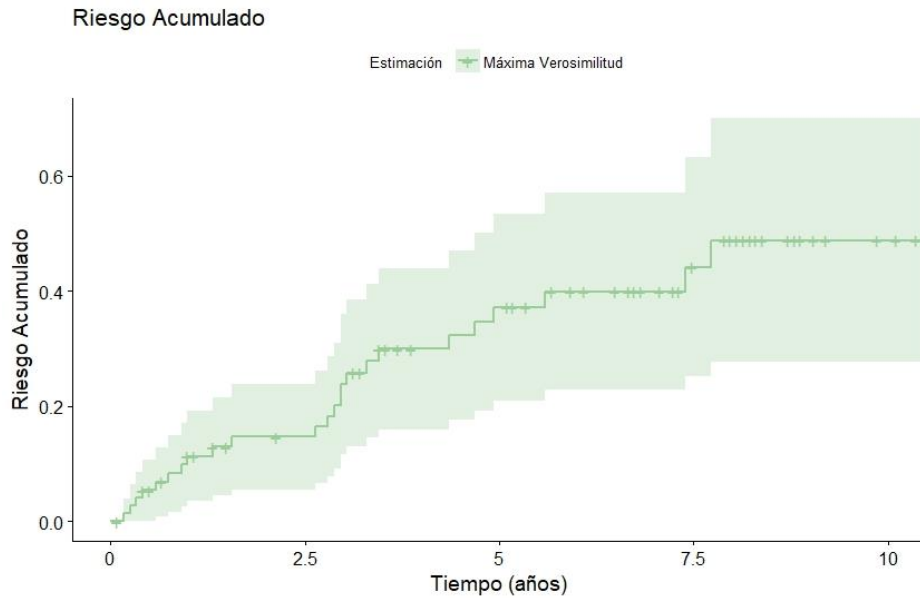


Figura 6 Función de riesgo acumulado global

### 2.4.2 Comparaciones por grupos

El método de Kaplan-Maier permite realizar comparaciones entre grupos de una misma variable, tal como se ha expuesto en el epígrafe **2.1.3** del proyecto.

Para este proyecto se ha considerado que, de entre los factores que se dispone en la base de datos, es interesante evaluar por grupos: Hepatocarcinoma, Edad, Sexo, Etiología, Tratamientos y Recidiva.

Por tanto, para cada variable de interés se construye un modelo con la función *survfit()* y el objeto *Surv* (“Longevidad” y “Muerte”) relacionándolo con el dicho factor del cuál queremos comparar los grupos. Con ello se obtiene una gráfica que ilustra los conjuntos, utilizando *ggsurvplot()* y se realiza el test Log-rank con la función *survdif()*. Este test contrasta las siguientes hipótesis:

$$H_0: S_1(t) = S_2(t) = \dots = S_k(t)$$

$$H_1: S_1(t) \neq S_2(t) \neq \dots \neq S_k(t)$$

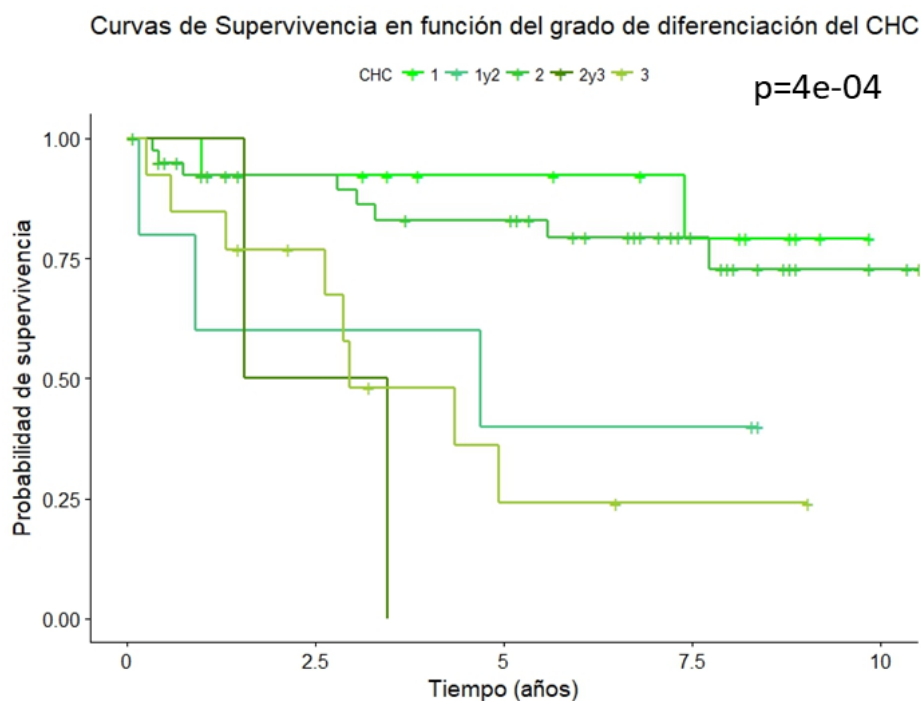
La hipótesis nula afirma que no existen diferencias estadísticamente significativas entre las funciones. Por contraposición, la hipótesis alternativa postula que sí las hay.

La cirrosis no es una variable interesante a evaluar por grupos porque prácticamente todos los individuos la padecen. Por otra parte, tampoco se considera evaluar esta comparación por tipo de cirugía ya que cada una se aplica según el grado de desarrollo de la enfermedad y por tanto no son comparables.

### 2.4.2.1 Hepatocarcinoma

Las variables que evalúan el grado de desarrollo del carcinoma hepático de los pacientes son “estadio de diferenciación del CHC” y “estadio tumoral”.

Se han creado las curvas de supervivencia en función del estadio de diferenciación del hepatocarcinoma. En la *Figura 7* se pueden observar las mencionadas curvas.



*Figura 7 Curvas de supervivencia en función de la diferenciación del CHC*

Se puede observar en la gráfica que, conforme aumenta el grado de diferenciación del cáncer, disminuye significativamente su probabilidad de supervivencia, para un nivel de significación del 5%. Ello se debe a que a mayor diferenciación, más avanzada se detectó la enfermedad, por lo que, según estas funciones, cuanto más desarrollada se opera menor posibilidad de supervivencia tienen los individuos (Sociedad Española de Oncología Médica, 2018, <https://www.seom.org/es/info-sobre-el-cancer/higado?showall=1>).

Los estadios que se consideran intermedios tienen un pronóstico peor que el del rango en el que en teoría se incluyen. Esto puede deberse a que es un método un tanto subjetivo, por lo que está íntimamente ligado con el profesional que lo dictamine.

En la *Figura 8* se evalúan de forma gráfica las funciones de supervivencia según la clasificación del estadio tumoral desarrollada por la AJCC.

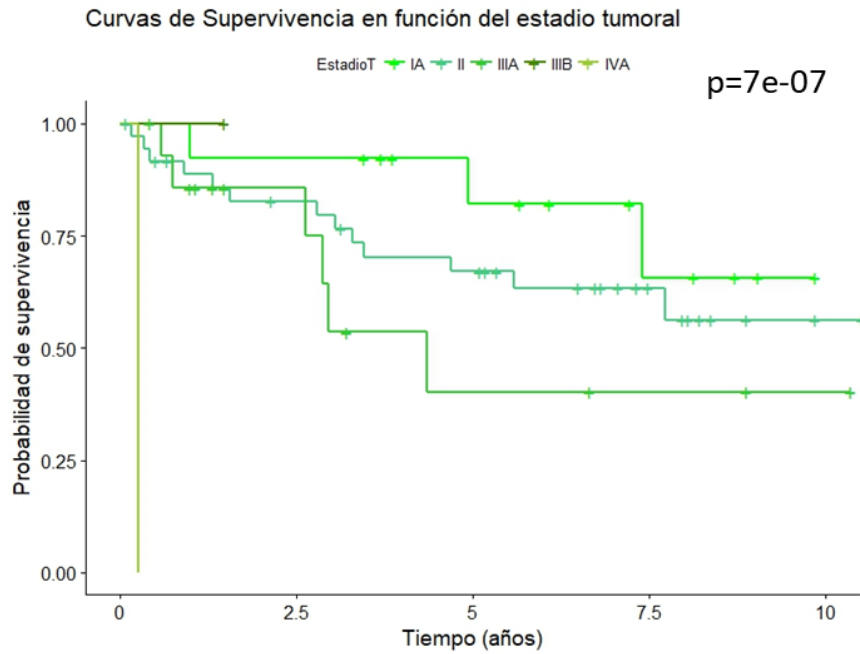


Figura 8 Curvas de supervivencia en función del estadio tumoral

En este caso, se observa la misma tendencia que con la variable de diferenciación histológica: a mayor gravedad de la enfermedad, menores posibilidades de superarla pese a la intervención quirúrgica. Los pacientes con el estadio tumoral más desarrollado presentan según la Figura 7 una probabilidad de supervivencia prácticamente nula, mientras que aquellos que presentan el estadio IA se quedan en una probabilidad cercana al 70%.

#### 2.4.2.2 Sexo

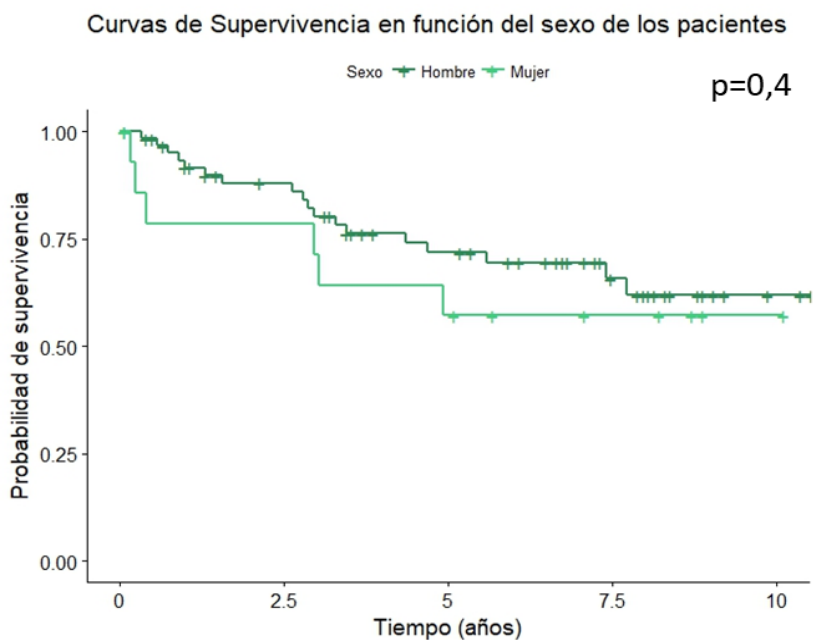


Figura 9 Curvas de supervivencia en función del sexo

La *Figura 9* muestra el resultado gráfico de las funciones de supervivencia para hombres y mujeres.

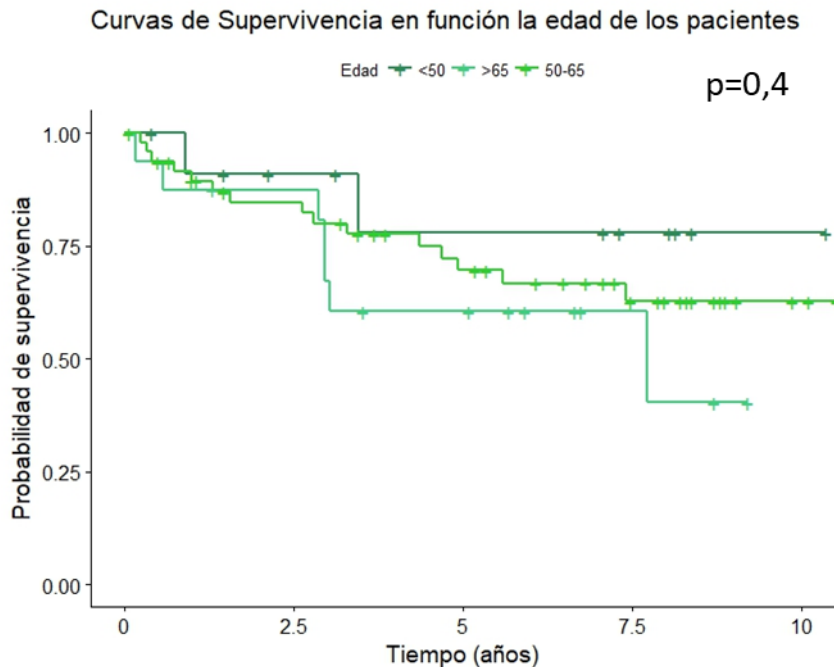
Utilizando el test log-Rank, se ha observado que las curvas no son significativamente distintas, para  $\alpha=0.05$ . Las mujeres que padecen de hepatocarcinoma tienen una probabilidad menos favorable según la *Figura 9*, no existen diferencias estadísticas entre ellas y el grupo de varones que padecen esta enfermedad.

### 2.4.2.3 Edad

La variable “Edad” es una variable continua por lo que previamente se ha dividido en grupos para poder contrastarla. Se han decidido tres rangos de edad en función del cómputo global de los pacientes:

- “<50” menos de cincuenta años
- “50-65” entre cincuenta y sesentaicinco años
- “>65” mayor de sesentaicinco años

Una vez definidos los tres grupos, se han creado las funciones de supervivencia para cada uno de ellos y en la *Figura 10* se observan las curvas de estas.

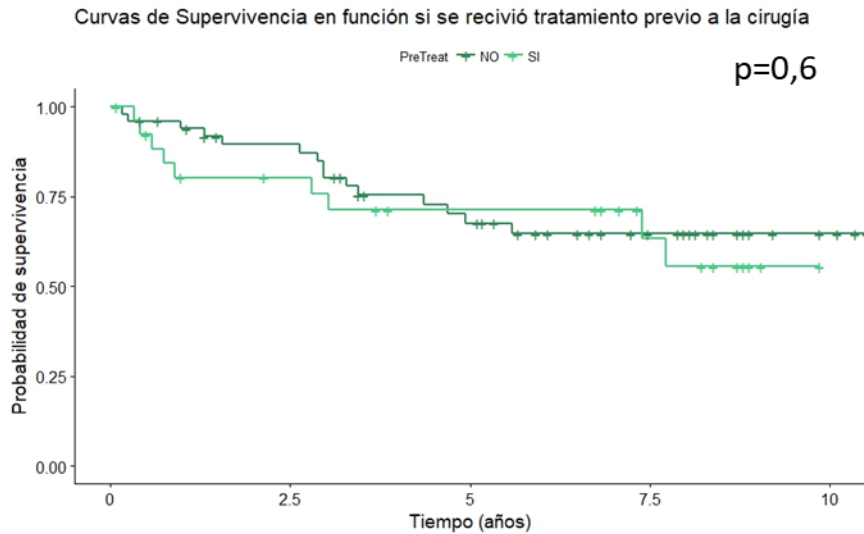


*Figura 10* Curvas de supervivencia en función de la edad

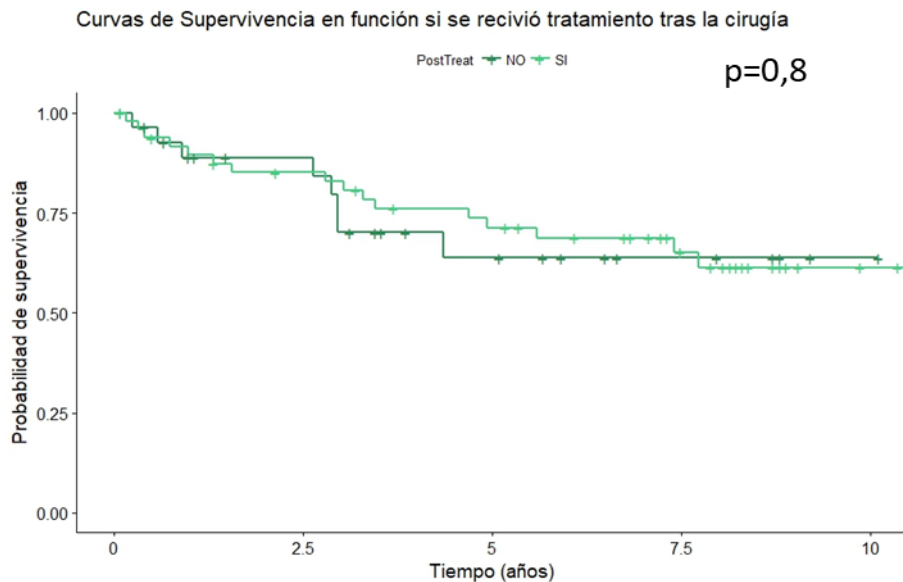
No existen diferencias significativas para los tres grupos. Aunque el Instituto Nacional del Cáncer sitúa la edad como el factor de riesgo más importante de cáncer en general (Instituto Nacional del Cáncer, 2018, <https://www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/edad>), en esta base de datos no se distribuyen con funciones de supervivencia estadísticamente distintas.

#### 2.4.2.4 Tratamientos

Se han realizado las funciones de supervivencia, separándolas en grupos conforme a si recibieron tratamientos o no, tanto antes como después de la cirugía, *Figura 11 y Figura 12*.



*Figura 11 Curvas de supervivencia en función de si se recibió tratamiento previo a la cirugía*



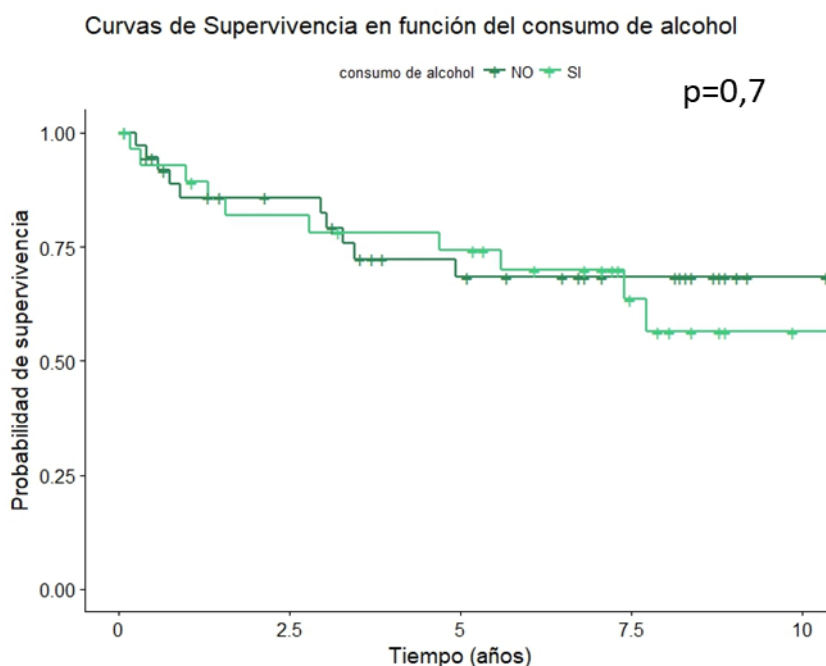
*Figura 12 Curvas de supervivencia en función de si se recibió tratamiento tras a la cirugía*

En ambos casos las curvas no han expresado una diferencia significativa ( $p\text{-valor} > 0.05$ ). Por tanto, los tratamientos (más allá de la cirugía) no tienen relevancia estadística en la probabilidad de sobrevivir a la enfermedad para este grupo de pacientes.

### 2.4.2.5 Etiología

Respecto a la etiología causante del CHC, se dividió en dos factores para analizarlos de forma individual: consumo elevado de alcohol e infección vírica.

Al realizar el contraste (log-Rank) para comparar los grupos de pacientes que consumían alcohol y los que no, éste no resultó estadísticamente significativo, tal como podemos ver en la *Figura 13*.



*Figura 13 Curvas de supervivencia en función del consumo de alcohol*

Las estadísticas globales indican el consumo de alcohol como un factor de riesgo muy relevante que aumenta el riesgo de padecer cáncer (Instituto Nacional del Cáncer, 2018, <https://www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/alcohol>), pero una vez desarrollado el análisis no existen diferencias entre las probabilidades de supervivencia de aquellos que lo desarrollaron por haber o no consumido alcohol.

En el caso de la infección vírica, las funciones de supervivencia entre las personas que las padecían (una o más) y las que no (expuestas en la *Figura 14*), tampoco son significativamente distintas. Los pacientes que padecían el virus de la hepatitis (ya sea B o C) no tienen probabilidad de supervivencia estadísticamente distinta que los que no la padecían.

El factor de riesgo más elevado para el CHC es la infección con virus de la hepatitis B y C (RUIZ, Eloy, et al., 2007), pero una vez detectado e intervenido, los pacientes no presentan distintas posibilidades de superar la enfermedad, que aquellos que la desarrollaron por otras causas.

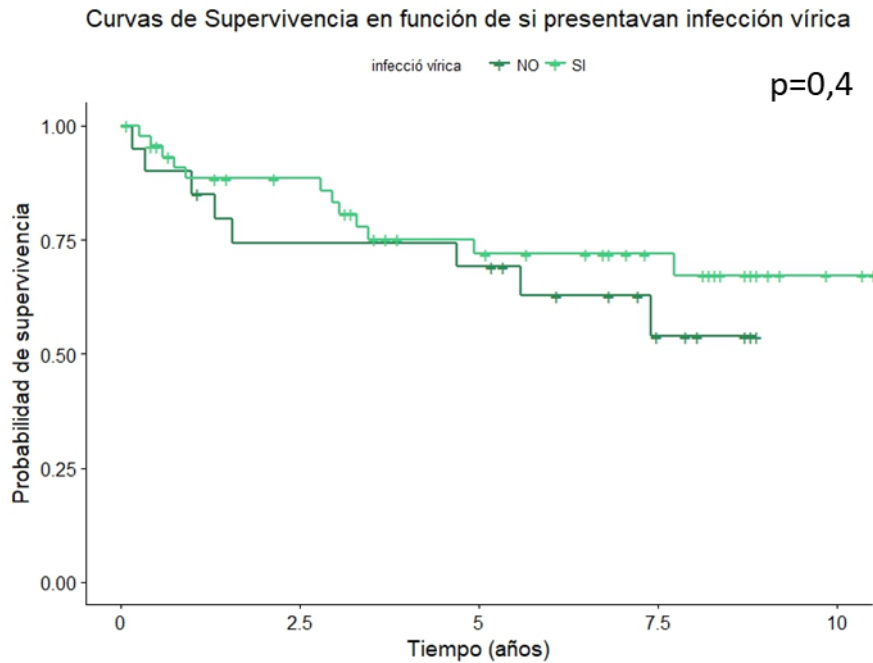


Figura 14 Curvas de supervivencia en función de si los pacientes presentaban infección vírica

#### 2.4.2.6 Recidiva

En la *Figura 15* se puede ver de forma clara que la curva de supervivencia de aquellos pacientes que presentaron recidiva tiene una pendiente mucho más acentuada que aquellos que no. Por tanto, como es lógico, los pacientes que sufren una reaparición del CHC tienen una probabilidad de supervivencia más baja. En este estudio la reaparición de la neoplasia se acentúa a partir de los 3 años de la operación.

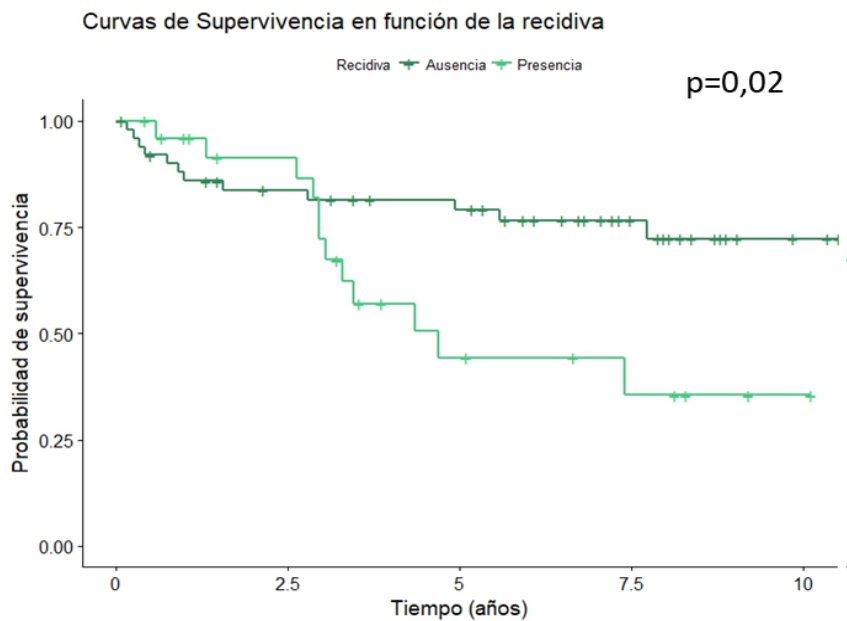


Figura 15 Curvas de supervivencia en función de la recidiva



Con la prueba Log-rank se obtuvo un p-valor inferior a 0.05, por lo que podemos asumir que existen diferencias estadísticas significativas entre las probabilidades de supervivencia de los pacientes que sufren recidiva y los que no.

### 2.4.3 Modelo de riesgos proporcionales de Cox

El modelo de Cox es utilizado para detectar relaciones existentes entre el riesgo que se produce en un determinado individuo en el estudio y algunas variables independientes y/o explicativas. En este apartado se lleva a cabo un primer análisis univariante de cada variable y uno multivariante de todas ellas.

#### 2.4.3.1 Univariante

La función utilizada en el programa estadístico R para estudiar un modelo de regresión de Cox es *coxph()* que está implementada en el paquete de supervivencia “*survival*”. A partir de las variables “Longevidad” y “Muerte” se ha creado un modelo de Cox para cada factor, los p-valores de las cuales se exponen en la *Tabla 7*.

*Tabla 7 p-valores de los modelos de Cox para cada variable*

Variable		p-valor
Sexo		0.44
Edad		0.08
Etanol		0.74
Virología		0.4
EstadioT	EstadioTII	0.32117
	EstadioTIIIA	0.10139
	EstadioTIIIB	0.99455
	EstadioTIVA	0.00142 **
PreTreat		0.65
PostTreat		0.81
HCC	HCC1 y 2	0.0666
	HCC2	0.573
	HCC2 y 3	0.12
	HCC3	0.12
Cirrosis		0.32

En la *Tabla 7* podemos ver que ninguna variable es significativa por sí sola, a excepción del estadio tumoral TIVA. Es decir, ninguna de ellas tiene relación en la supervivencia de los pacientes de forma individual (a excepción de la ya comentada).

### 2.4.3.2 Multivariante

Para poder analizar las variables de forma conjunta, se realiza un modelo multivariante de Cox con pasos hacia delante (SOLANES, Ingrid, et al., 2007). Se construye dicho modelo con la misma función *coxph()* pero esta vez relacionando el objeto *Surv* con todos los factores a la vez.

	coef	exp(coef)	se(coef)	z	p
HCC1 y 2	1.8916	6.6301	1.0779	1.75	0.079
HCC2	0.4241	1.5283	0.8673	0.49	0.625
HCC2 y 3	2.9912	19.9097	1.2460	2.40	0.016
HCC3	1.6922	5.4316	1.0671	1.59	0.113
Edad	0.0789	1.0821	0.0381	2.07	0.039
SexoMujer	0.5650	1.7594	0.8051	0.70	0.483
EtanolSI	-0.0137	0.9864	0.8414	-0.02	0.987
VirologiaSI	-1.1106	0.3294	0.7737	-1.44	0.151
EstadioTII	0.9193	2.5075	0.8359	1.10	0.271
EstadioTIIIA	1.0459	2.8461	1.1069	0.94	0.345
EstadioTIIIB	0.0000	1.0000	0.0000	NA	NA
EstadioTIVA	2.1312	8.4248	3.8151	0.56	0.576
RecidivaSI	0.8415	2.3198	0.5978	1.41	0.159
PreTreatSI	1.4847	4.4135	0.7028	2.11	0.035
PostTreatSI	0.2171	1.2425	0.9516	0.23	0.820

Muchas de las variables no son significativas por lo que no aportan información al modelo. Por esa razón, se han ido eliminando variables una a una (cada vez, la que presentaba un p-valor mayor), hasta dar con el siguiente modelo en el que todas contribuyen de forma significativa.

	coef	exp(coef)	se(coef)	z	p
HCC1 y 2	2.4386	11.4574	0.9756	2.50	0.0124
HCC2	0.3541	1.4249	0.7928	0.45	0.6551
HCC2 y 3	3.4975	33.0338	1.0908	3.21	0.0013
HCC3	2.1181	8.3154	0.8066	2.63	0.0086
Edad	0.0829	1.0864	0.0323	2.57	0.0103
PreTreatSI	0.9180	2.5042	0.4940	1.86	0.0632

La variable HCC no se ha eliminado ya que uno de sus grupos si aporta información al modelo.

Para analizar este modelo, se realiza un resumen del mismo a través de la orden *summary()*.

	coef	exp(coef)	se(coef)	z	Pr(> z )
HCC1 y 2	2.43863	11.45735	0.97561	2.500	0.01243 *
HCC2	0.35411	1.42490	0.79283	0.447	0.65514
HCC2 y 3	3.49753	33.03378	1.09081	3.206	0.00134 **
HCC3	2.11811	8.31538	0.80658	2.626	0.00864 **
Edad	0.08291	1.08645	0.03231	2.567	0.01027 *
PreTreatSI	0.91796	2.50418	0.49404	1.858	0.06316 .

---

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

		exp(coef)	exp(-coef)	lower	.95	upper	.95
HCC1	y 2	11.457	0.08728	1.6930			77.540
HCC2		1.425	0.70180	0.3013			6.740
HCC2	y 3	33.034	0.03027	3.8946			280.193
HCC3		8.315	0.12026	1.7113			40.406
Edad		1.086	0.92043	1.0198			1.157
PreTreatSI		2.504	0.39933	0.9509			6.595

Concordance= 0.775 (se = 0.064 )  
 Rsquare= 0.271 (max possible= 0.91 )  
 Likelihood ratio test= 23.66 on 6 df, p=6e-04  
 Wald test = 21.93 on 6 df, p=0.001  
 Score (logrank) test = 26.59 on 6 df, p=2e-04

El test de razón de verosimilitudes, test de Wald y test score presentan p-valores menores a 0.05, por lo que asumimos que el modelo es explicativo de forma estadísticamente significativa. La calidad de esta explicación no es buena dado que el modelo presenta una  $R^2$  con un valor muy bajo.

Para contrastar si se cumple la hipótesis de riesgos proporcionales, ejecutamos la función *cox.zph()* que se basa en el test de Kolmogorov, que permite determinar la bondad del ajuste de las distribuciones de probabilidad.

		rho	chisq		p
HCC1	y	2	-0.1779	0.8396	0.360
HCC2			-0.0728	0.1190	0.730
HCC2	y	3	0.1100	0.2971	0.586
HCC3			-0.0287	0.0185	0.892
Edad			-0.0394	0.0367	0.848
PreTreatSI			-0.2083	1.0209	0.312
GLOBAL		NA	4.2905	0.637	

Todas las variables presentan un p-valor superior a 0.05, por lo que no rechazamos la hipótesis de riesgos proporcionales.

Utilizando la función *ggcoxdiagnostics()* se grafican los residuos del modelo, mediante DFBeta, y se muestran en la *Figura 10*. Esta es la estimación de mínimos cuadrados que se utiliza para detectar específicamente la influencia que una observación ejerce en la estimación de los parámetros del modelo (MOSCOSO, José Alfredo Jiménez, et al., 200).

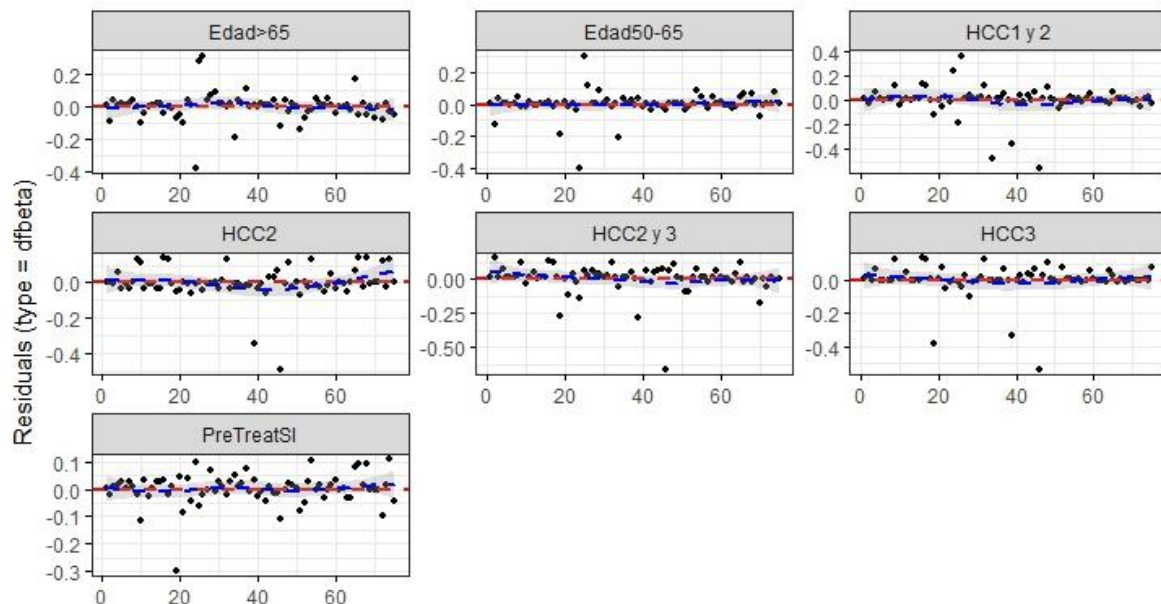


Figura 16 Residuos de los parametros mediante DFBeta

En la *Figura 16* no se aprecia ninguna tendencia en los residuos de las variables, por lo que se acepta que los residuos son independientes del tiempo de supervivencia y el modelo es correcto.

#### 2.4.4 Estimación paramétrica de la función supervivencia

Los métodos de estimación paramétricos suponen una tendencia continua que cuando se dispone de datos censurados no se puede asumir (LEE, Elisa, et al. 1997).

El modelo Weibull de regresión es uno de los métodos paramétricos más populares. Proporciona una estimación de la función de riesgo, así como los coeficientes para las covariables. Este tipo de modelo presenta ciertas dificultades técnicas, y por ello rara vez se utiliza en la literatura médica en comparación con el modelo de riesgo proporcional semiparamétrico (ZHANG, Zhongheng, 2016).

GU, Xiangdong, et al., (2014), desarrolla un paquete de R *straweib*, para ajustar un modelo de Weibull estratificado, apropiado para los resultados censurados por intervalo. En el caso de este proyecto, hay una gran cantidad de datos censurados por la derecha, por lo que, tal como se menciona en el apartado **2.1.2**, no sería apropiado utilizar este método.

### 3. Conclusiones

En el presente proyecto se ha evaluado la supervivencia de un grupo de pacientes, a los que se les realizó un seguimiento tras ser intervenidos de cáncer hepático.

Las primeras conclusiones que se pueden extraer son que los pacientes analizados cumplen con las tendencias de los registros existentes sobre la mencionada enfermedad. Es decir, la mayoría de pacientes son varones, con edades comprendidas entre los 40 y 80 años, la etiología de mayor relevancia es el VHC y prácticamente todos presentan cirrosis.

Los pacientes operados para tratar un hepatocarcinoma no presentan una tasa de mortandad elevada, gracias a que han podido ser intervenidos en una fase poco desarrollada de la enfermedad. Cabe destacar, que la clave radica en haber realizado las operaciones en esa fase de la enfermedad.

Al comparar los pacientes por grupos se puede observar que tan solo aquellas variables que clasifican la gravedad de la enfermedad presentan funciones de supervivencia significativamente distintas desde un punto de vista estadístico (“Diferenciación del CHC”, “Estadio tumoral” y “recidiva”). En otras palabras, cuanto más desarrollado se interviene el CHC, menor supervivencia (de forma estadísticamente significativa) tienen los pacientes.

Sin embargo, a partir del origen del cáncer hepático, el sexo y la edad de los pacientes o si se recibieron tratamientos, antes o después de la cirugía, las funciones de supervivencia de los individuos no resultan variar de forma significativa.

Este estudio también ha permitido ver que ninguna de las variables afecta de forma directa a la supervivencia de los pacientes mediante los modelos univariantes de Cox. Al realizar un modelo multivariante, con todas ellas en conjunto inicialmente, ha permitido ver que las variables “Edad”, “Pretratamiento” y “Diferenciación del CHC”, sí que construyen un modelo significativamente explicativo. Pero éste presenta un valor de  $R^2$  que indica que es un tanto pobre describiendo la supervivencia de los pacientes.

## 4. Glosario

Cirrosis: Enfermedad crónica del hígado, en la cual las células hepáticas son reemplazadas por tejido cicatricial.

Hepatitis: Inflamación del hígado

Hepatocarcinoma: Tumor maligno que se desarrolla en el hígado cuando las células de éste comienzan a crecer de forma desmesurada

Nódulo: Crecimiento o masa que puede ser maligna (cancerosa) o benigna.

Recidiva: Cáncer que ha reaparecido tras un período de no detección.

Tumor: Masa anormal de tejido que resulta cuando las células se multiplican de forma descontrolada.

VHB (Virus de la Hepatitis B): Es un ente vírico que causa infección hepática (entre otros que también pueden causar hepatitis). La hepatitis B es una enfermedad infecciosa vírica prevenible (existe una vacuna específica para la hepatitis B).

VHC (Virus de la Hepatitis C): Es un virus que causa hepatitis. Ésta es una enfermedad infecciosa de la que no se dispone vacuna.

## 5. Bibliografía

ARRIBALZAGA, Eduardo (2007). <<Interpretación de las curvas de supervivencia>>. Revista chilena de cirugía.

BRENNER, Hermann, et al., (2006). <<Up-to-date and precise estimates of cancer patient survival: model-based period analysis>>. American journal of epidemiology.

FERNÁNDEZ, Pita (1995). <<Análisis de supervivencia>>. Unidad de Epidemiología Clínica.

GU, Xiangdong, et al, (2014). <<Stratified Weibull regression model for interval-censored data>>. The R journal.

GIJBELS, Irène, et al., (1993). << Strong representations of the survival function estimator for truncated and censored data with applications>>. Journal of Multivariate Analysis.

GIJBELS, Irène, (2010). <<Censored data>>. Wiley Interdisciplinary Reviews: Computational Statistics.

GRAMATGES, Anissa, (2002). <<Aplicación y técnicas del análisis de supervivencia en las investigaciones clínicas>>. Revista Cubana de Hematología, Inmunología y Hemoterapia.

HOSMER, David, et al., (2012). << Regression Models for Survival Data>>. Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition, 2008.

KLEINBAUM, David, et al., (2010). << Survival analysis>>. New York: Springer.

KLEIN, John P, et al.,(2005). <<Survival analysis: techniques for censored and truncated data>>. Springer Science & Business Media.

LEE, Elisa T.; GO, Oscar T, (1997). <<Survival analysis in public health research>>. Annual review of public health.

MURILLO, Adrián, et al., 2013. <<Resección hepática por hepatocarcinoma: estudio comparativo entre pacientes menores y mayores de 70 años>>.

MOSCOSO, José Alfredo Jiménez; SUARÉZ, Luis Francisco Rincón, (2002). <<Una generalización de la estadística DFBeta>>. Revista Colombiana de Estadística.

REBASA, Pere (2005). <<Conceptos básicos del análisis de supervivencia>>. Cirugía Española.

RUIZ, Eloy, et al., (2007). << Resultados a corto y largo plazo de la Resección Hepática por Hepatocarcinoma>>. Análisis de 232 Resecciones Consecutivas. Revista de Gastroenterología del Perú.

SOLANES, Ingrid, et al., (2013). <<Factores de riesgo de mortalidad en la EPOC>>. Archivos de Bronconeumología.

VARELA, María, et al., (2010). <<Tratamiento del carcinoma hepatocelular en España. Análisis de 705 casos en 62 centros>>. Medicina clínica.

ZHANG, Zhongheng, (2016). <<Parametric regression model for survival data: Weibull regression model as an example>>. Annals of translational medicine.

Associació catalana de Malalts d'Hepatitis, 08/05/2018, <http://asscat-hepatitis.org/hepatocarcinoma/>

Hospital Universitario Ramón y Cajal, 2018, [http://www.hrc.es/bioest/Supervivencia\\_7.html](http://www.hrc.es/bioest/Supervivencia_7.html)

Instituto Nacional del cáncer, 2018, <https://www.cancer.gov/espanol/publicaciones/diccionario/def/quimioembolizacion>

Instituto Nacional del cáncer, 2018, <https://www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/edad>

Sociedad Española de Oncología Médica, 2018, <https://www.seom.org/es/info-sobre-el-cancer/higado?showall=1>



## 6. Anexos

### 6.1 Estadios tumorales según la clasificación de la AJCC

Etapa AJCC	TNM	Descripción
IA	T1a N0 M0	Un solo tumor de 2 cm (4/5 de pulgada) o de menor tamaño que no ha crecido hacia los vasos sanguíneos (T1a)
IB	T1b N0 M0	Un solo tumor que mide más de 2cm (4/5 de pulgada) y que no ha crecido hacia los vasos sanguíneos (T1b). El cáncer no se ha propagado a los ganglios linfáticos adyacentes (N0) ni a sitios distantes (M0).
II	T2 N0 M0	Un solo tumor que mide más de 2cm (4/5 de pulgada) y que ha crecido hacia los vasos sanguíneos, O más de un tumor sin que ninguno mida más de 5 cm (aproximadamente dos pulgadas) de ancho (T2). No se ha propagado a los ganglios linfáticos adyacentes (N0), ni a sitios distantes (M0).
IIIA		Más de un tumor, con al menos uno que mide más de 5 cm de ancho (T3). No se ha propagado a los ganglios linfáticos adyacentes (N0), ni a sitios distantes (M0).
IIIB	T4 N0 M0	Hay al menos un tumor (de cualquier tamaño) que ha crecido hacia una rama principal de una vena grande del hígado (vena hepática o la vena porta) (T4). No se ha propagado a los ganglios linfáticos adyacentes (N0), ni a sitios distantes (M0).
IVA	Cualquier T N1 M0	Un solo tumor o muchos tumores de cualquier tamaño (Cualquier T) que se ha propagado a los ganglios linfáticos adyacentes (N1), pero no a sitios distantes (M0).
IVB	Cualquier T cualquier N M1	Un solo tumor o muchos tumores de cualquier tamaño (Cualquier T). Puede o no haberse propagado a los ganglios linfáticos adyacentes (cualquier N). El cáncer se ha propagado a órganos distantes como los huesos o los pulmones (M1).

## 6.2 Código de análisis compilado por Rmarkdown

# Importación del conjunto de datos

Los datos se encuentran en un fichero con con extensión .xlsx, por ese motivo se emplea la función “read\_excel” de la librería readxl.

```
library(readxl)
Dataset_HCC <- read_excel("C:/Users/Usuario/Desktop/UOC/TFM/Dataset_HC
C.xlsx",
  sheet = "Hoja2")
#Se muestran las primeras filas para verificar la importaación.
head(Dataset_HCC,5)

## # A tibble: 5 x 14
##   HCC Etiologia Etanol Virologia Cirrosis EstadioT Cirugia
Muerte
##   <chr> <chr>      <dbl>      <dbl>      <dbl> <chr>      <chr>
<dbl>
## 1 1      VHC          0          1          1 II      Hepatectomía ~
0
## 2 <NA> <NA>          NA          NA          NA <NA>      Hepatectomía ~
NA
## 3 2 y 3 VHC          0          1          1 II      Hepatectomía ~
1
## 4 2      VHC          0          1          1 IIIA     Hepatectomía ~
0
## 5 1      VHB          0          1          1 <NA>      Tumorectomía
0
## # ... with 6 more variables: Longevidad <dbl>, Recidiva <dbl>,
## # PostTreat <dbl>, PreTreat <dbl>, Edad <dbl>, Sexo <dbl>
```

## Recodificación de los datos

Se comprueba mediante la función “str” el tipo de variables de los datos y se recodifican como factores las variables categóricas.

```
Dataset_HCC$HCC<-factor(Dataset_HCC$HCC)
Dataset_HCC$Etiologia<-factor(Dataset_HCC$Etiologia,levels = c("0","1"
), labels = c("NO","SI"))
Dataset_HCC$Cirrosis<-factor(Dataset_HCC$Cirrosis,levels = c("0","1"),
labels = c("NO","SI"))
Dataset_HCC$Cirugia<-factor(Dataset_HCC$Cirugia)
Dataset_HCC$EstadioT<-factor(Dataset_HCC$EstadioT)
Dataset_HCC$PreTreat<-factor(Dataset_HCC$PreTreat,levels = c("0","1"),
labels = c("NO","SI"))
Dataset_HCC$PostTreat<-factor(Dataset_HCC$PostTreat,levels = c("0","1"
), labels = c("NO","SI"))
Dataset_HCC$Recidiva<-factor(Dataset_HCC$Recidiva,levels = c("0","1"),
labels = c("NO","SI"))
Dataset_HCC$Sexo<-factor(Dataset_HCC$Sexo,levels = c("0","1"), labels
= c("Hombre","Mujer"))
Dataset_HCC$Etanol<-factor(Dataset_HCC$Etanol,levels = c("0","1"), lab
els = c("NO","SI"))
```

```

Dataset_HCC$Virologia<-factor(Dataset_HCC$Virologia,levels = c("0","1"
), labels = c("NO","SI"))
str(Dataset_HCC)

## Classes 'tbl_df', 'tbl' and 'data.frame':   83 obs. of  14 variables:
## $ HCC      : Factor w/ 5 levels "1","1 y 2","2",...: 1 NA 4 3 1 3
3 3 NA 5 ...
## $ Etiologia : Factor w/ 2 levels "NO","SI": NA NA NA NA NA NA NA NA
NA NA NA ...
## $ Etanol   : Factor w/ 2 levels "NO","SI": 1 NA 1 1 1 2 NA 2 1 1
...
## $ Virologia : Factor w/ 2 levels "NO","SI": 2 NA 2 2 2 1 NA 2 2 2
...
## $ Cirrosis : Factor w/ 2 levels "NO","SI": 2 NA 2 2 2 2 2 2 2 2 .
..
## $ EstadioT : Factor w/ 5 levels "IA","II","IIIA",...: 2 NA 2 3 NA
3 4 2 NA 3 ...
## $ Cirugia  : Factor w/ 25 levels "Hemipatectomía izquierda",...: 7
2 7 7 24 7 7 7 12 16 ...
## $ Muerte   : num  0 NA 1 0 0 0 0 0 0 0 ...
## $ Longevidad: num  3.12 NA 3.45 10.36 8.79...
## $ Recidiva  : Factor w/ 2 levels "NO","SI": 1 NA 2 1 1 1 1 1 2 2 .
..
## $ PostTreat : Factor w/ 2 levels "NO","SI": 1 1 2 2 1 2 1 2 1 1 ..
.
## $ PreTreat  : Factor w/ 2 levels "NO","SI": 1 1 1 1 1 1 1 1 1 1 ..
.
## $ Edad     : num  46 55 43 48 54 61 65 53 68 49 ...
## $ Sexo     : Factor w/ 2 levels "Hombre","Mujer": 1 1 1 1 1 1 1 1 1
1 1 ...

```

## Estadística Descriptiva

```

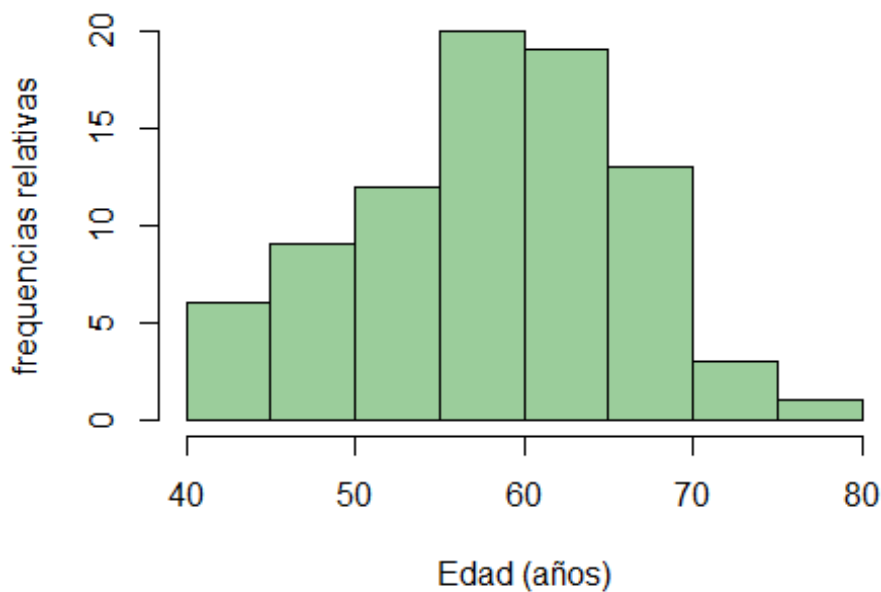
#Edad
summary(Dataset_HCC$Edad)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  40.00  54.00   59.00   58.58  64.00   80.00

hist(Dataset_HCC$Edad, ylab="frecuencias relativas", xlab = "Edad (años)",
main="Edad de los pacientes", col="darkseagreen3" )

```

## Edad de los pacientes



```
#Sexo
prop.table(table(Dataset_HCC$Sexo))

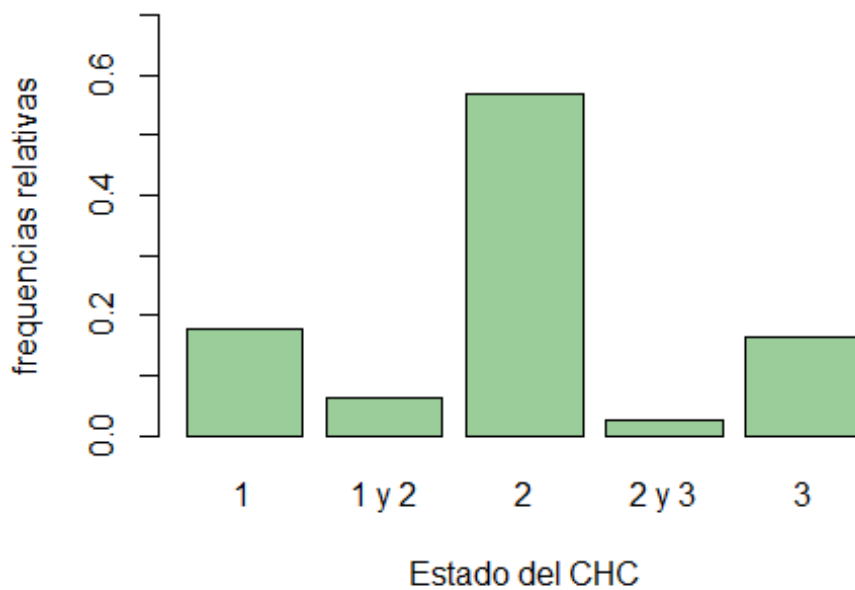
##
##   Hombre   Mujer
## 0.8072289 0.1927711

#Estado del HCC
prop.table(table(Dataset_HCC$HCC))

##
##      1      1 y 2      2      2 y 3      3
## 0.17721519 0.06329114 0.56962025 0.02531646 0.16455696

barplot(prop.table(table(Dataset_HCC$HCC)), xlab= "Estado del CHC" , y
lab="frecuencias relativas", col="darkseagreen3", main = "Estado del H
CC de los pacientes" ,ylim= c(0,0.7))
```

## Estado del HCC de los pacientes



```
##(1 = Bien diferenciado,2 = Moderadamente diferenciado,3 = Poco diferenciado)
```

```
#Etiologia
```

```
length(grep("VHB", Dataset_HCC$Etiologia, ignore.case=TRUE))
```

```
## [1] 0
```

```
length(grep("VHC", Dataset_HCC$Etiologia, ignore.case=TRUE))
```

```
## [1] 0
```

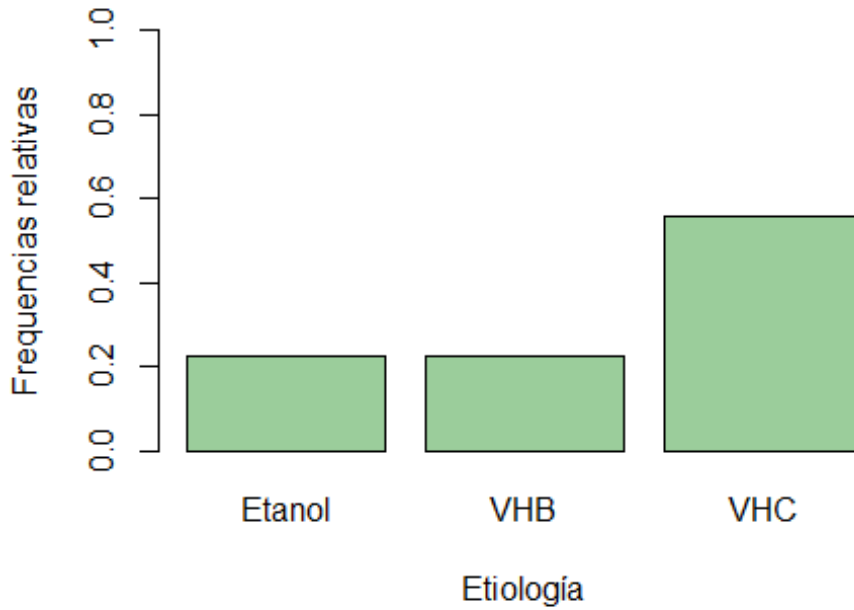
```
length(grep("VHB", Dataset_HCC$Etiologia, ignore.case=TRUE))
```

```
## [1] 0
```

```
df.Etiologia<-data.frame(Etiologia=c(rep("VHB",14),rep("VHC",35),rep("Etanol",14)))
```

```
barplot(prop.table(table(df.Etiologia)),xlab="Etiología", ylab="Frecuencias relativas", col="darkseagreen3", main ="Etiología de los CHC",ylim= c(0,1))
```

## Etiología de los CHC



```
#Cirrosis
prop.table(table(Dataset_HCC$Cirrosis))

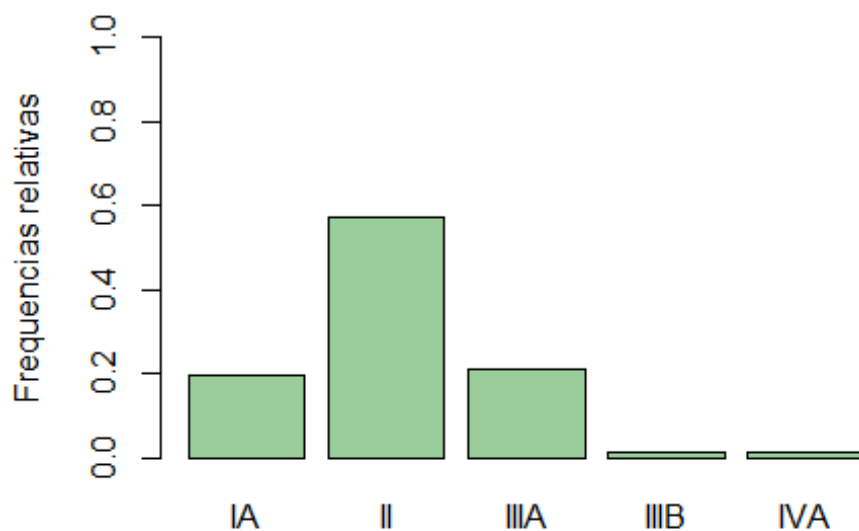
##
##      NO      SI
## 0.1025641 0.8974359

#Estadio tumoral
prop.table(table(Dataset_HCC$EstadioT))

##
##      IA      II      IIIA      IIIB      IVA
## 0.19444444 0.56944444 0.20833333 0.01388889 0.01388889

barplot(prop.table(table(Dataset_HCC$EstadioT)) , ylab="Frecuencias re
lativas", col="darkseagreen3", main = "Estadio tumoral de los CHC",yli
m= c(0,1))
```

## Estadio tumoral de los CHC



```
#Cirugia
prop.table(table(Dataset_HCC$Cirugia))

##
##           Hemipatectomía izquierda
##                0.01204819
##           Hepatectomía derecha
##                0.02409639
##           Hepatectomía izquierda
##                0.01204819
## Hepatectomía izquierda y segmentectomía
##                0.01204819
##           Hepatectomía parcial
##                0.01204819
##           Hepatectomía segmento VI
##                0.01204819
##           Hepatectomía TOH
##                0.65060241
##           Hepatectomía total TOH
##                0.01204819
## Lobectomía derecha y segmentectomía
##                0.01204819
##           Pieza de trasplante
##                0.01204819
##           Resección
##                0.01204819
##           Resección atípica segmento IV
##                0.01204819
##           Resección en cuña
##                0.01204819
##           Resección segmento IV
```

```

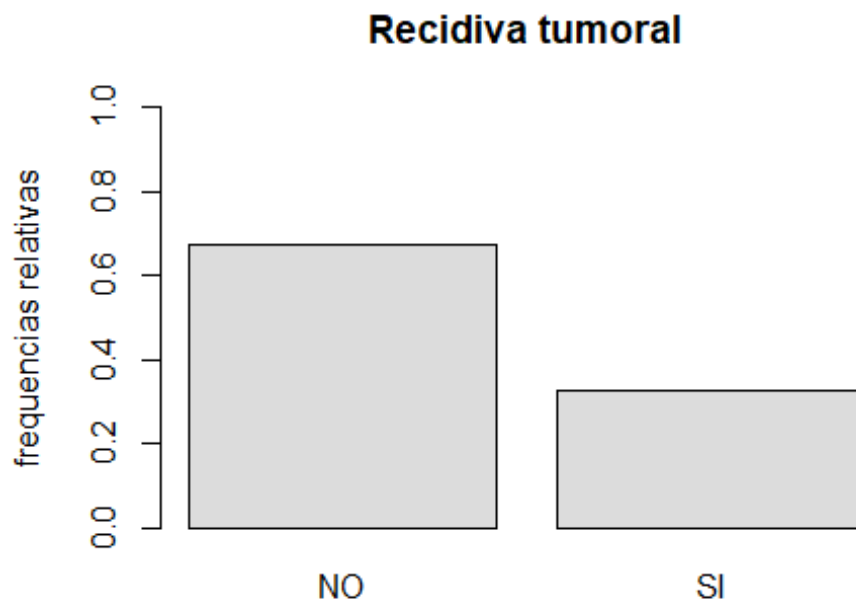
##          0.01204819
##      Resección tumoral
##          0.01204819
##      Segmentectomía
##          0.02409639
##      Segmentectomía II
##          0.01204819
##      Segmentectomía III
##          0.01204819
##      Segmentectomía IV
##          0.01204819
##      Segmentectomía lateral izquierda
##          0.02409639
##      Segmentectomía VII y VIII
##          0.02409639
##      Segmentectomía VIVII
##          0.01204819
##      Segmentectomía VVIII
##          0.01204819
##      Tumorectomía
##          0.02409639
##      Tumorectomía segmento VIII
##          0.01204819

```

```
##Hepatectomía TOH => 0.65%
```

```
#Recidiva
```

```
barplot(prop.table(table(Dataset_HCC$Recidiva)) , ylab="frecuencias re
lativas", col="gainsboro", main = "Recidiva tumoral",ylim= c(0,1))
```





```

#Pre-tratamiento y Post-tratamiento
prop.table(table(Dataset_HCC$PreTreat ))

##
##          NO          SI
## 0.6626506 0.3373494

prop.table(table(Dataset_HCC$PostTreat ))

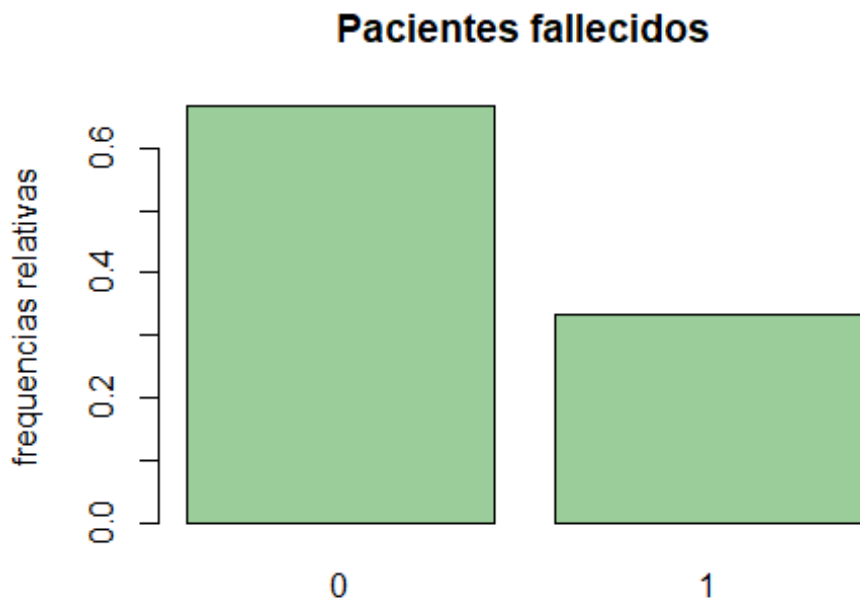
##
##          NO          SI
## 0.4096386 0.5903614

#Pacientes fallecidos
prop.table(table(Dataset_HCC$Muerte))

##
##          0          1
## 0.6666667 0.3333333

barplot(prop.table(table(Dataset_HCC$Muerte)), ylab= "frecuencias rela
tivas", col = "darkseagreen3", main = "Pacientes fallecidos")

```



```

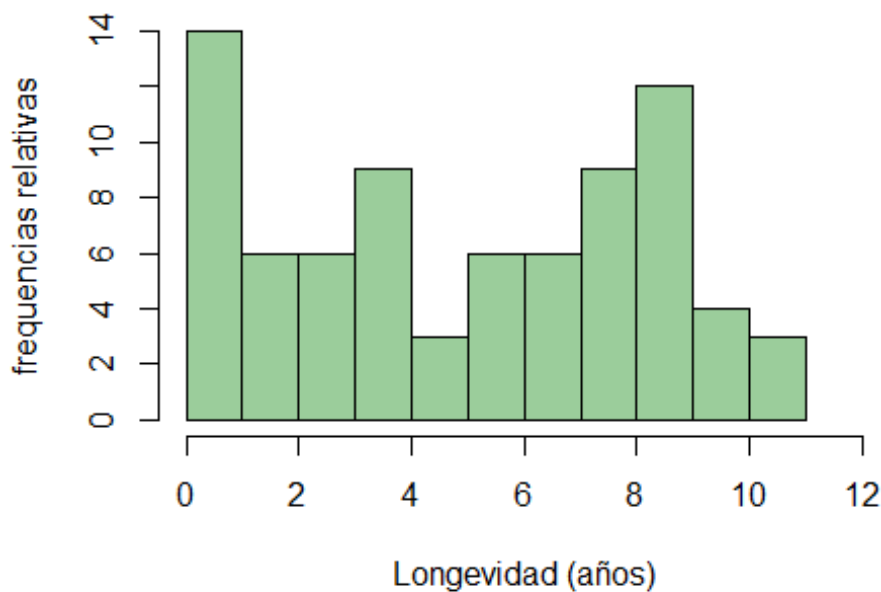
#Longevidad (años)
summary(Dataset_HCC$Longevidad)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.08219 1.70548 5.13699 4.96206 7.95205 10.52055    5

hist(Dataset_HCC$Longevidad, ylab="frecuencias relativas", xlab = "Lon
gevidad (años)", main="Longevidad de los pacientes", col="darkseagreen
3",xlim= c(0,12) )

```

## Longevidad de los pacientes



## Estimación de la función supervivencia

```
library(survminer)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

```
## Loading required package: magrittr
```

```
library(survival)
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
#kaplan-meier
```

```
Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
```

```
## [1] 3.12328767+ NA? 3.45205479 10.35616438+ 8.7945205
```

```
5+
```

```
## [6] 8.87671233+ 1.47945205+ 10.52054795+ 3.53424658+ 1.4794520
```

```
5+
```

```
## [11] 10.10958904+ 9.86301370+ 5.58904110 5.91780822+ 4.9315068
```

```

5
## [16] 1.06849315+ 9.86301370+ 7.97260274+          NA 9.2054794
5+
## [21] 8.87671233+ 1.31506849+          NA 9.04109589+ 2.9589041
1
## [26] 8.71232877+ 0.24657534 3.20547945+ 8.71232877+ 0.9041095
9
## [31] 8.38356164+ 0.16438356 2.95890411 6.49315068+ 2.1369863
0+
## [36] 8.38356164+ 8.13698630+ 8.21917808+ 7.89041096+ 8.3013698
6+
## [41] 8.05479452+ 7.47945205+ 0.57534247 0.98630137+ 7.3972602
7
## [46] 3.69863014+ 4.68493151 8.79452055+ 3.45205479+ 7.7260274
0
## [51] 3.86301370+ 0.98630137 2.87671233 5.67123288+ 7.2328767
1+
## [56] 0.41095890+ 6.73972603+ 1.31506849 5.09589041+ 2.7945205
5
## [61] 7.06849315+ 5.17808219+ 7.06849315+ 7.31506849+ 6.8219178
1+
## [66] 6.65753425+ 0.08219178+ 0.65753425+ 6.82191781+          N
A
## [71] 4.35616438 3.04109589 0.41095890 6.08219178+ 0.3287671
2
## [76] 0.49315068+ 1.56164384          NA? 0.08219178+ 3.2876712
3
## [81] 5.34246575+ 0.73972603 2.63013699

km <- survfit(Surv(Longevidad, Muerte) ~ 1, data =Dataset_HCC, type =
"kaplan-meier")
km

## Call: survfit(formula = Surv(Longevidad, Muerte) ~ 1, data = Datasets_HCC,
type = "kaplan-meier")
##
## 5 observations deleted due to missingness
##      n events median 0.95LCL 0.95UCL
## 78.00 24.00      NA 7.73      NA

summary(km)

## Call: survfit(formula = Surv(Longevidad, Muerte) ~ 1, data = Datasets_HCC,
type = "kaplan-meier")
##
## 5 observations deleted due to missingness
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164     76      1    0.987 0.0131  0.962 1.000
## 0.247     75      1    0.974 0.0184  0.938 1.000
## 0.329     74      1    0.961 0.0223  0.918 1.000
## 0.411     73      1    0.947 0.0256  0.898 0.999
## 0.575     70      1    0.934 0.0286  0.879 0.992
## 0.740     68      1    0.920 0.0313  0.861 0.984

```

```
## 0.904 67 1 0.906 0.0337 0.843 0.975
## 0.986 66 1 0.893 0.0359 0.825 0.966
## 1.315 63 1 0.878 0.0380 0.807 0.956
## 1.562 59 1 0.864 0.0402 0.788 0.946
## 2.630 57 1 0.848 0.0422 0.770 0.935
## 2.795 56 1 0.833 0.0441 0.751 0.924
## 2.877 55 1 0.818 0.0458 0.733 0.913
## 2.959 54 2 0.788 0.0489 0.698 0.890
## 3.041 52 1 0.773 0.0502 0.680 0.878
## 3.288 49 1 0.757 0.0516 0.662 0.865
## 3.452 48 1 0.741 0.0529 0.644 0.852
## 4.356 43 1 0.724 0.0544 0.625 0.839
## 4.685 42 1 0.707 0.0558 0.605 0.825
## 4.932 41 1 0.689 0.0570 0.586 0.811
## 5.589 37 1 0.671 0.0584 0.565 0.796
## 7.397 24 1 0.643 0.0623 0.532 0.777
## 7.726 22 1 0.614 0.0660 0.497 0.758
```

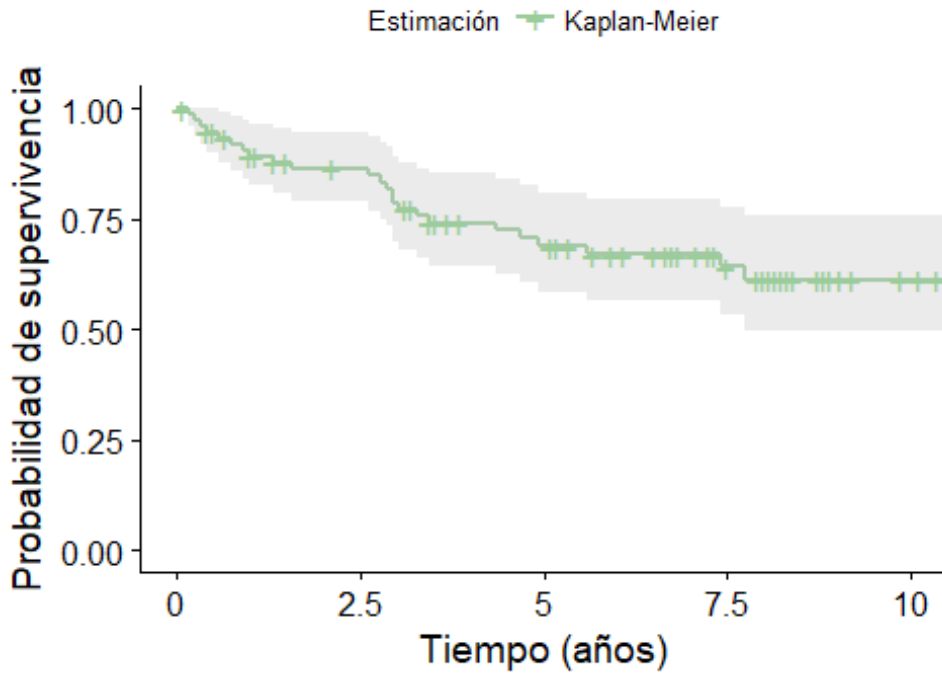
```
print(km)
```

```
## Call: survfit(formula = Surv(Longevidad, Muerte) ~ 1, data = Dataset_HCC,
## type = "kaplan-meier")
##
## 5 observations deleted due to missingness
## n events median 0.95LCL 0.95UCL
## 78.00 24.00 NA 7.73 NA
```

```
#curva de supervivencia
```

```
ggsurvplot(fit = km, data = Dataset_HCC, conf.int = T, title = "Curva
de Supervivencia",
xlab = "Tiempo (años)", ylab = "Probabilidad de supervivencia", legend.title = "Estimación",
legend.labs = "Kaplan-Meier", palette = "darkseagreen3")
```

## Curva de Supervivencia



```
#Estimación del riesgo acumulado (Máxima Verosimilitud)
kmR <- survfit(Surv(Longevidad, Muerte) ~ 1, Dataset_HCC)
summary(kmR)
```

```
## Call: survfit(formula = Surv(Longevidad, Muerte) ~ 1, data = Dataset_HCC)
```

```
##
```

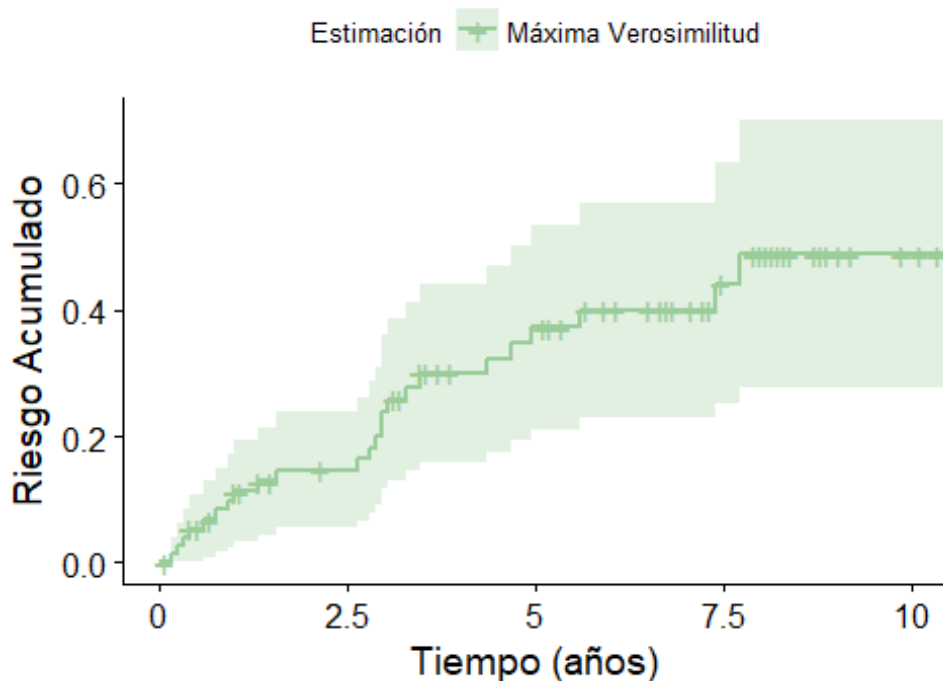
```
## 5 observations deleted due to missingness
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	0.164	76	1	0.987	0.0131	0.962	1.000
##	0.247	75	1	0.974	0.0184	0.938	1.000
##	0.329	74	1	0.961	0.0223	0.918	1.000
##	0.411	73	1	0.947	0.0256	0.898	0.999
##	0.575	70	1	0.934	0.0286	0.879	0.992
##	0.740	68	1	0.920	0.0313	0.861	0.984
##	0.904	67	1	0.906	0.0337	0.843	0.975
##	0.986	66	1	0.893	0.0359	0.825	0.966
##	1.315	63	1	0.878	0.0380	0.807	0.956
##	1.562	59	1	0.864	0.0402	0.788	0.946
##	2.630	57	1	0.848	0.0422	0.770	0.935
##	2.795	56	1	0.833	0.0441	0.751	0.924
##	2.877	55	1	0.818	0.0458	0.733	0.913
##	2.959	54	2	0.788	0.0489	0.698	0.890
##	3.041	52	1	0.773	0.0502	0.680	0.878
##	3.288	49	1	0.757	0.0516	0.662	0.865
##	3.452	48	1	0.741	0.0529	0.644	0.852
##	4.356	43	1	0.724	0.0544	0.625	0.839
##	4.685	42	1	0.707	0.0558	0.605	0.825
##	4.932	41	1	0.689	0.0570	0.586	0.811
##	5.589	37	1	0.671	0.0584	0.565	0.796

```
## 7.397      24      1      0.643  0.0623      0.532      0.777
## 7.726      22      1      0.614  0.0660      0.497      0.758
```

```
ggsurvplot(kmR, fun = "cumhaz", xlab = "Tiempo (años)", censor = T, ylab = "Riesgo Acumulado", title = "Riesgo Acumulado", legend.title = "Estimación", legend.labs = "Máxima Verosimilitud", palette = "darkseagreen3")
```

## Riesgo Acumulado



## Diferencias entre grupos

```
library(survminer)
library(survival)
library(ggplot2)
library(dplyr)
#Hepatocarcinoma
km0=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ HCC, data = Dataset_HCC)
summary(km0)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
```

```
## HCC, data = Dataset_HCC)
```

```
##
```

```
## 8 observations deleted due to missingness
```

```
## HCC=1
```

```
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
## 0.986 13 1 0.923 0.0739 0.789 1
```

```
## 7.397 7 1 0.791 0.1375 0.563 1
```

```
##
```

```
## HCC=1 y 2
```

```
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
## 0.164 5 1 0.8 0.179 0.516 1
```

```

## 0.904      4      1      0.6  0.219      0.293      1
## 4.685      3      1      0.4  0.219      0.137      1
##
##                      HCC=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.329    40      1    0.975  0.0247    0.928    1.000
## 0.411    39      1    0.950  0.0345    0.885    1.000
## 0.740    35      1    0.923  0.0429    0.843    1.000
## 2.795    30      1    0.892  0.0513    0.797    0.999
## 3.041    29      1    0.861  0.0580    0.755    0.983
## 3.288    28      1    0.831  0.0636    0.715    0.965
## 5.589    23      1    0.794  0.0703    0.668    0.945
## 7.726    12      1    0.728  0.0904    0.571    0.929
##
##                      HCC=2 y 3
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 1.56     2      1     0.5   0.354    0.125     1
## 3.45     1      1     0.0   NaN      NA        NA
##
##                      HCC=3
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.247    13      1    0.923  0.0739    0.7890    1.000
## 0.575    12      1    0.846  0.1001    0.6711    1.000
## 1.315    11      1    0.769  0.1169    0.5711    1.000
## 2.630     8      1    0.673  0.1362    0.4527    1.000
## 2.877     7      1    0.577  0.1468    0.3504    0.950
## 2.959     6      1    0.481  0.1506    0.2602    0.888
## 4.356     4      1    0.361  0.1536    0.1565    0.831
## 4.932     3      1    0.240  0.1418    0.0756    0.764

```

```
print(km0)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
##      HCC, data = Dataset_HCC)
##
##      8 observations deleted due to missingness
##              n events median 0.95LCL 0.95UCL
## HCC=1      13      2      NA      NA      NA
## HCC=1 y 2   5      3   4.68   0.904     NA
## HCC=2      42      8      NA      NA      NA
## HCC=2 y 3   2      2   2.51   1.562     NA
## HCC=3      13      8   2.96   2.630     NA

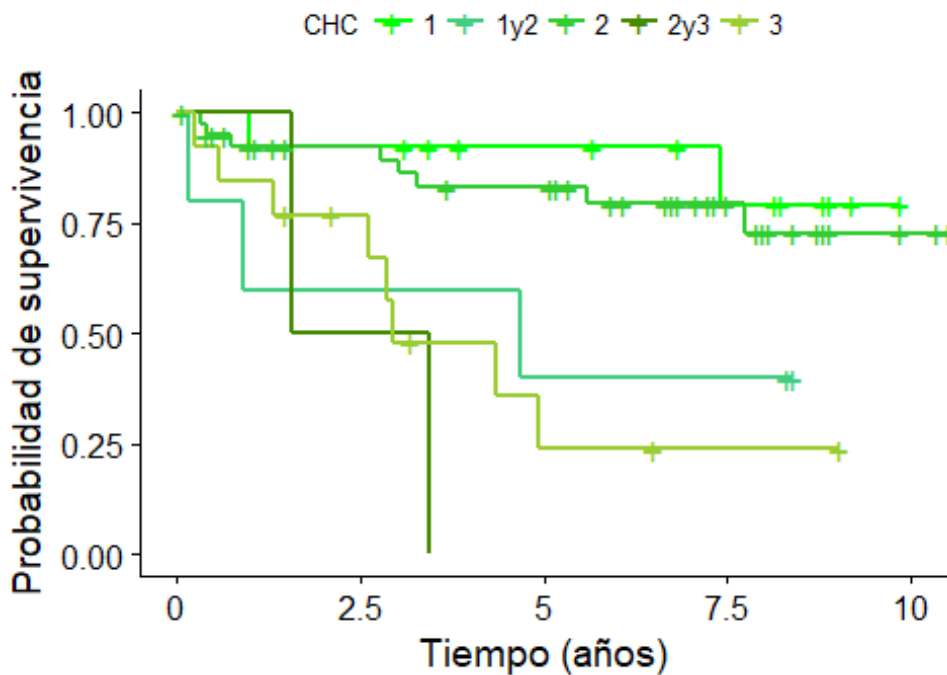
```

```

gg survplot(km0, data = Dataset_HCC, title = "Curvas de Supervivencia en
función del grado de diferenciación del CHC", xlab="Tiempo (años)", yla
b="Probabilidad de supervivencia", main="KM Plot by Sexo", conf.int =
F, legend.labs = c("1", "1y2", "2", "2y3", "3"), legend.title = "CHC", pale
tte = c("green", "seagreen3", "limegreen", "chartreuse4", "yellowgreen"))

```

## Curvas de Supervivencia en función del grado



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ HCC , data = Dataset_HCC)
```

```
## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
## ~
## HCC, data = Dataset_HCC)
##
## n=75, 8 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## HCC=1     13         2      5.072      1.86      2.40
## HCC=1 y 2  5         3      1.423      1.75      1.87
## HCC=2     42         8     13.129      2.00      4.69
## HCC=2 y 3  2         2      0.444      5.46      5.61
## HCC=3     13         8      2.933      8.76     10.24
##
## Chisq= 20.3 on 4 degrees of freedom, p= 4e-04
```

```
km1=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ EstadioT , data = Dataset_HCC )
```

```
summary(km1)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
## EstadioT, data = Dataset_HCC)
##
## 15 observations deleted due to missingness
##           EstadioT=IA
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.986     13         1   0.923  0.0739      0.789      1
```



```

## 4.932      9      1    0.821  0.1169      0.621      1
## 7.397      5      1    0.656  0.1740      0.390      1
##
##          EstadioT=II
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164   36     1    0.972  0.0274    0.920    1.000
## 0.329   35     1    0.944  0.0382    0.873    1.000
## 0.411   34     1    0.917  0.0461    0.831    1.000
## 0.904   31     1    0.887  0.0532    0.789    0.998
## 1.315   30     1    0.858  0.0591    0.749    0.982
## 1.562   29     1    0.828  0.0640    0.712    0.963
## 2.795   27     1    0.797  0.0686    0.674    0.944
## 3.041   26     1    0.767  0.0725    0.637    0.923
## 3.288   24     1    0.735  0.0762    0.600    0.900
## 3.452   23     1    0.703  0.0793    0.563    0.877
## 4.685   22     1    0.671  0.0819    0.528    0.852
## 5.589   18     1    0.634  0.0854    0.486    0.825
## 7.726    9     1    0.563  0.1008    0.396    0.800
##
##          EstadioT=IIIA
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.575   14     1    0.929  0.0688    0.803    1.000
## 0.740   13     1    0.857  0.0935    0.692    1.000
## 2.630    8     1    0.750  0.1294    0.535    1.000
## 2.877    7     1    0.643  0.1488    0.408    1.000
## 2.959    6     1    0.536  0.1579    0.301    0.955
## 4.356    4     1    0.402  0.1658    0.179    0.902
##
##          EstadioT=IIIB
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##
##          EstadioT=IVA
##  time      n.risk      n.event      survival      std.err
## 0.247      1.000      1.000      0.000      NaN
## lower 95% CI upper 95% CI
##          NA          NA

```

```
print(km1)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
```

```
## EstadioT, data = Dataset_HCC)
```

```
##
```

```
## 15 observations deleted due to missingness
```

```
##          n events median 0.95LCL 0.95UCL
```

```
## EstadioT=IA    13      3    NA    7.40    NA
```

```
## EstadioT=II   38     13    NA    5.59    NA
```

```
## EstadioT=IIIA 15      6  4.356  2.88    NA
```

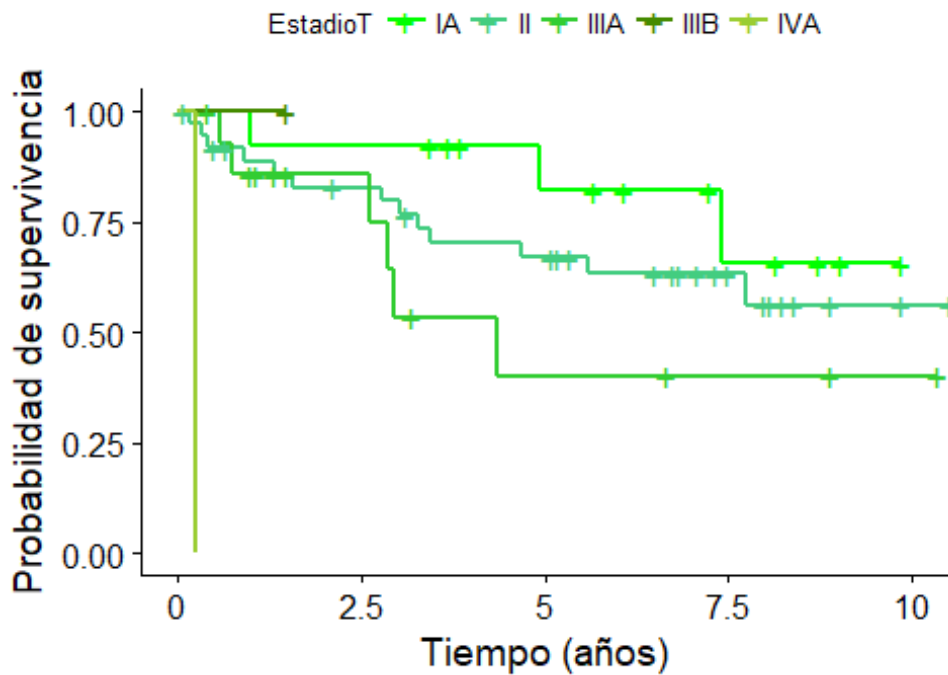
```
## EstadioT=IIIB  1      0    NA    NA     NA
```

```
## EstadioT=IVA   1      1  0.247  NA     NA
```

```
ggsurvplot(km1, data = Dataset_HCC, title = "Curvas de Supervivencia en
función del estadio tumoral", xlab="Tiempo (años)", ylab="Probabilidad
de supervivencia", main="KM Plot by Sexo", conf.int = F, legend.labs
```

```
= c("IA", "II", "IIIA", "IIIB", "IVA"), legend.title = "EstadioT", palette = c("green", "seagreen3", "limegreen", "chartreuse4", "yellowgreen"))
```

## Curvas de Supervivencia en función del estadio



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ EstadioT , data = Dataset_HCC)
```

```
## Call:
## survdif(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
## ~ EstadioT, data = Dataset_HCC)
##
## n=68, 15 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## EstadioT=IA  13         3   5.7523   1.31692   1.7659
## EstadioT=II  38        13  13.3212   0.00774   0.0185
## EstadioT=IIIA 15         6   3.7457   1.35665   1.6544
## EstadioT=IIIB  1         0   0.1502   0.15021   0.1528
## EstadioT=IVA   1         1   0.0305  30.77863  31.2559
##
## Chisq= 34.2 on 4 degrees of freedom, p= 7e-07
```

```
#Sexo
km2=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Sexo ,
data = Dataset_HCC )
summary(km2)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
## Sexo, data = Dataset_HCC)
##
```

```

## 5 observations deleted due to missingness
##
##           Sexo=Hombre
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.329   62     1    0.984  0.0160    0.953    1.000
## 0.575   59     1    0.967  0.0228    0.923    1.000
## 0.740   57     1    0.950  0.0280    0.897    1.000
## 0.904   56     1    0.933  0.0323    0.872    0.999
## 0.986   55     1    0.916  0.0359    0.849    0.989
## 1.315   52     1    0.899  0.0393    0.825    0.979
## 1.562   48     1    0.880  0.0427    0.800    0.968
## 2.630   46     1    0.861  0.0458    0.776    0.956
## 2.795   45     1    0.842  0.0486    0.752    0.943
## 2.877   44     1    0.823  0.0512    0.728    0.929
## 2.959   43     1    0.803  0.0534    0.705    0.915
## 3.288   40     1    0.783  0.0557    0.681    0.901
## 3.452   39     1    0.763  0.0578    0.658    0.885
## 4.356   34     1    0.741  0.0603    0.632    0.869
## 4.685   33     1    0.718  0.0625    0.606    0.852
## 5.589   30     1    0.694  0.0649    0.578    0.834
## 7.397   19     1    0.658  0.0710    0.532    0.813
## 7.726   17     1    0.619  0.0767    0.486    0.789
##
##           Sexo=Mujer
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164   14     1    0.929  0.0688    0.803    1.000
## 0.247   13     1    0.857  0.0935    0.692    1.000
## 0.411   12     1    0.786  0.1097    0.598    1.000
## 2.959   11     1    0.714  0.1207    0.513    0.995
## 3.041   10     1    0.643  0.1281    0.435    0.950
## 4.932    9     1    0.571  0.1323    0.363    0.899

```

```
print(km2)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
##   Sexo, data = Dataset_HCC)
##
##   5 observations deleted due to missingness
##           n events median 0.95LCL 0.95UCL
## Sexo=Hombre 63     18    NA     7.73     NA
## Sexo=Mujer  15     6     NA     3.04     NA

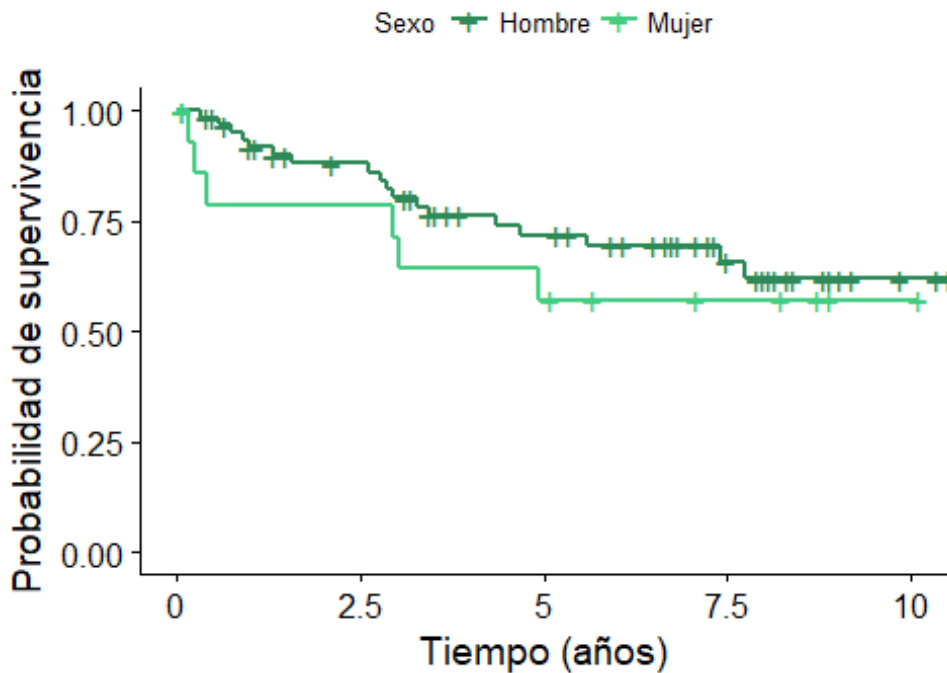
```

```

ggsurvplot(km2, data = Dataset_HCC, title = "Curvas de Supervivencia en
función del sexo de los pacientes", xlab="Tiempo (años)", ylab="Probabi
lidad de supervivencia", main="KM Plot by Sexo", conf.int = F, legend.
labs = c("Hombre", "Mujer"), legend.title = "Sexo", palette = c("seagre
en", "seagreen3"))

```

## Curvas de Supervivencia en función del sex



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Sexo ,
data = Dataset_HCC)

## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
~
##     Sexo, data = Dataset_HCC)
##
## n=78, 5 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Sexo=Hombre 63      18   19.48    0.112    0.597
## Sexo=Mujer  15       6    4.52    0.482    0.597
##
## Chisq= 0.6  on 1 degrees of freedom, p= 0.4

#Edad
Dataset_HCC$Edad[Dataset_HCC$Edad<50]<-"<50"
Dataset_HCC$Edad[Dataset_HCC$Edad>=50&Dataset_HCC$Edad<=65]<-"50-65"
Dataset_HCC$Edad[Dataset_HCC$Edad>65]<-">65"
km3=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Edad ,
data = Dataset_HCC )
summary(km3)

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
Edad, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
##           Edad=<50
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```

## 0.904      11      1    0.909  0.0867      0.754      1
## 3.452       7      1    0.779  0.1413      0.546      1
##
##                               Edad=>65
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164   16     1    0.938  0.0605    0.826    1.000
## 0.575   15     1    0.875  0.0827    0.727    1.000
## 2.877   13     1    0.808  0.1000    0.634    1.000
## 2.959   12     2    0.673  0.1204    0.474    0.956
## 3.041   10     1    0.606  0.1258    0.403    0.910
## 7.726    3     1    0.404  0.1850    0.165    0.991
##
##                               Edad=50-65
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.247   48     1    0.979  0.0206    0.940    1.000
## 0.329   47     1    0.958  0.0288    0.903    1.000
## 0.411   46     1    0.938  0.0349    0.871    1.000
## 0.740   43     1    0.916  0.0404    0.840    0.998
## 0.986   42     1    0.894  0.0449    0.810    0.986
## 1.315   39     1    0.871  0.0493    0.780    0.973
## 1.562   37     1    0.847  0.0533    0.749    0.959
## 2.630   36     1    0.824  0.0567    0.720    0.943
## 2.795   35     1    0.800  0.0598    0.691    0.927
## 3.288   33     1    0.776  0.0627    0.662    0.909
## 4.356   29     1    0.749  0.0660    0.631    0.891
## 4.685   28     1    0.723  0.0689    0.599    0.871
## 4.932   27     1    0.696  0.0713    0.569    0.851
## 5.589   24     1    0.667  0.0740    0.536    0.829
## 7.397   17     1    0.628  0.0794    0.490    0.804

```

```
print(km3)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
##      Edad, data = Dataset_HCC)
##
##      5 observations deleted due to missingness
##              n events median 0.95LCL 0.95UCL
## Edad=<50     12      2     NA      NA      NA
## Edad=>65     16      7    7.73    2.96     NA
## Edad=50-65  50     15     NA    7.40     NA

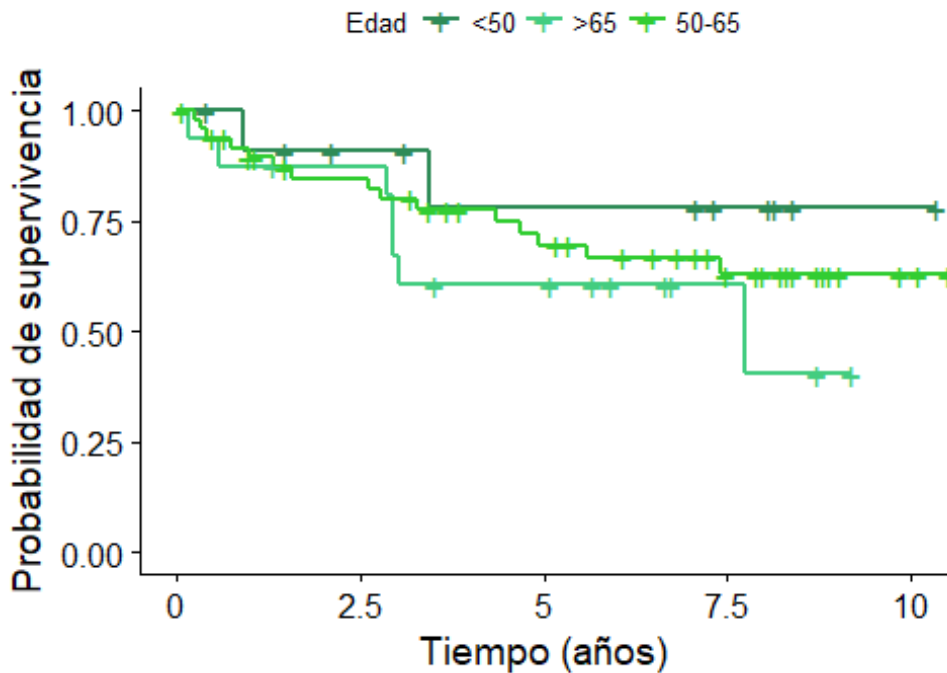
```

```

ggsurvplot(km3, data = Dataset_HCC, title = "Curvas de Supervivencia en
función la edad de los pacientes", xlab="Tiempo (años)", ylab="Probabil
idad de supervivencia", main="KM Plot by Edad", conf.int = F, legend.l
abs = c("<50", ">65", "50-65"), legend.title = "Edad", palette = c("seagr
een", "seagreen3", "limegreen"))

```

## Curvas de Supervivencia en función la edad



```
survdiff ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Edad ,
data = Dataset_HCC )

## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
~
##     Edad, data = Dataset_HCC)
##
## n=78, 5 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Edad=<50   12         2     3.70    0.7791    0.923
## Edad=>65   16         7     4.83    0.9758    1.229
## Edad=50-65 50        15    15.47    0.0145    0.041
##
## Chisq= 1.8  on 2 degrees of freedom, p= 0.4

#Tratamientos
km4=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ PreTreat
t , data = Dataset_HCC )
summary(km4)

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
PreTreat) ~
##     PreTreat, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
##           PreTreat=NO
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164    50     1    0.980  0.0198    0.942    1.000
## 0.247    49     1    0.960  0.0277    0.907    1.000
```

```
## 0.986 46 1 0.939 0.0341 0.875 1.000
## 1.315 44 1 0.918 0.0394 0.844 0.998
## 1.562 40 1 0.895 0.0446 0.812 0.987
## 2.630 39 1 0.872 0.0490 0.781 0.973
## 2.877 38 1 0.849 0.0528 0.751 0.959
## 2.959 37 2 0.803 0.0591 0.695 0.928
## 3.288 33 1 0.779 0.0621 0.666 0.911
## 3.452 32 1 0.754 0.0648 0.638 0.893
## 4.356 29 1 0.728 0.0676 0.607 0.874
## 4.685 28 1 0.702 0.0700 0.578 0.854
## 4.932 27 1 0.676 0.0721 0.549 0.833
## 5.589 23 1 0.647 0.0747 0.516 0.811
```

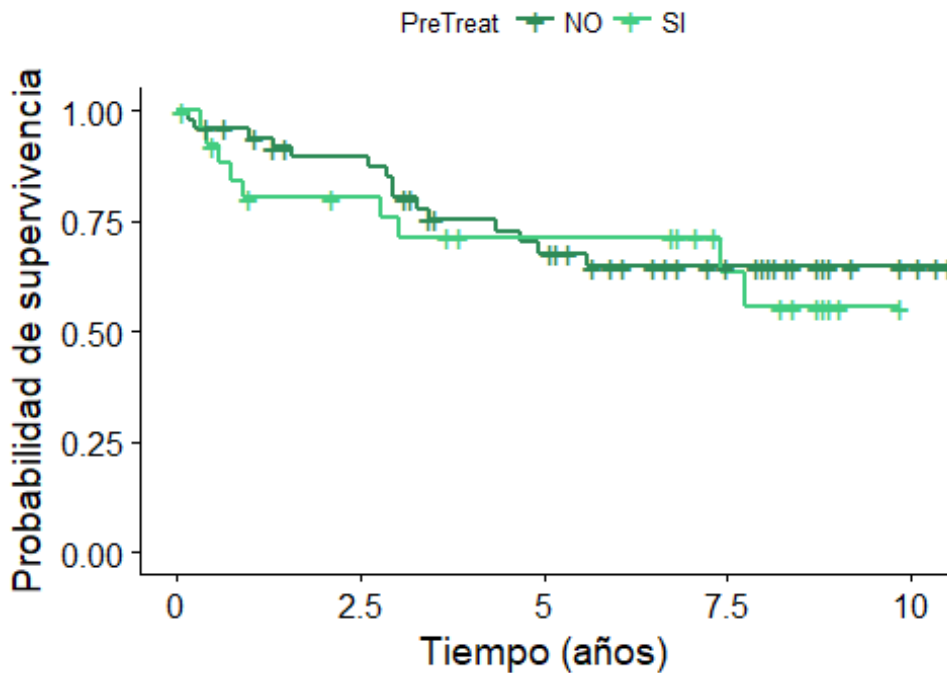
```
##
## PreTreat=SI
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.329 26 1 0.962 0.0377 0.890 1.000
## 0.411 25 1 0.923 0.0523 0.826 1.000
## 0.575 23 1 0.883 0.0636 0.767 1.000
## 0.740 22 1 0.843 0.0722 0.712 0.997
## 0.904 21 1 0.803 0.0792 0.662 0.974
## 2.795 18 1 0.758 0.0864 0.606 0.948
## 3.041 17 1 0.713 0.0921 0.554 0.919
## 7.397 9 1 0.634 0.1109 0.450 0.893
## 7.726 8 1 0.555 0.1221 0.361 0.854
```

```
print(km4)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
## PreTreat, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
## n events median 0.95LCL 0.95UCL
## PreTreat=NO 50 15 NA NA NA
## PreTreat=SI 28 9 NA 7.4 NA
```

```
ggsurvplot(km4, data = Dataset_HCC, title = "Curvas de Supervivencia en
función si se recibió tratamiento previo a la cirugía", xlab="Tiempo (a
ños)", ylab="Probabilidad de supervivencia", main="KM Plot byPreTreat"
, conf.int = F, legend.labs = c("NO", "SI"), legend.title = "PreTreat", p
alette = c("seagreen", "seagreen3"))
```

## Curvas de Supervivencia en función si se re



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ PreTreat , data = Dataset_HCC )
```

```
## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
## ~
## PreTreat, data = Dataset_HCC)
##
## n=78, 5 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## PreTreat=NO 50         15    16.05   0.0681   0.206
## PreTreat=SI 28          9     7.95   0.1373   0.206
##
## Chisq= 0.2  on 1 degrees of freedom, p= 0.6
```

```
km5=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ PostTreat , data = Dataset_HCC )
```

```
summary(km5)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
## PostTreat, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
##           PostTreat=NO
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.247    28     1    0.964  0.0351   0.898    1.000
## 0.575    26     1    0.927  0.0496   0.835    1.000
## 0.904    24     1    0.889  0.0607   0.777    1.000
## 2.630    19     1    0.842  0.0734   0.710    0.999
```



```

## 2.877 18 1 0.795 0.0829 0.648 0.975
## 2.959 17 2 0.701 0.0959 0.537 0.917
## 4.356 11 1 0.638 0.1063 0.460 0.884
##
##
## PostTreat=SI
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164 48 1 0.979 0.0206 0.940 1.000
## 0.329 47 1 0.958 0.0288 0.903 1.000
## 0.411 46 1 0.938 0.0349 0.871 1.000
## 0.740 44 1 0.916 0.0401 0.841 0.998
## 0.986 43 1 0.895 0.0445 0.812 0.986
## 1.315 42 1 0.874 0.0483 0.784 0.973
## 1.562 40 1 0.852 0.0518 0.756 0.959
## 2.795 38 1 0.829 0.0550 0.728 0.945
## 3.041 37 1 0.807 0.0579 0.701 0.929
## 3.288 35 1 0.784 0.0607 0.673 0.912
## 3.452 34 1 0.761 0.0631 0.647 0.895
## 4.685 32 1 0.737 0.0655 0.619 0.877
## 4.932 31 1 0.713 0.0676 0.592 0.859
## 5.589 28 1 0.688 0.0698 0.564 0.839
## 7.397 19 1 0.652 0.0749 0.520 0.816
## 7.726 17 1 0.613 0.0797 0.475 0.791

```

```
print(km5)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
## PostTreat, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
## n events median 0.95LCL 0.95UCL
## PostTreat=NO 29 8 NA 4.36 NA
## PostTreat=SI 49 16 NA 7.73 NA

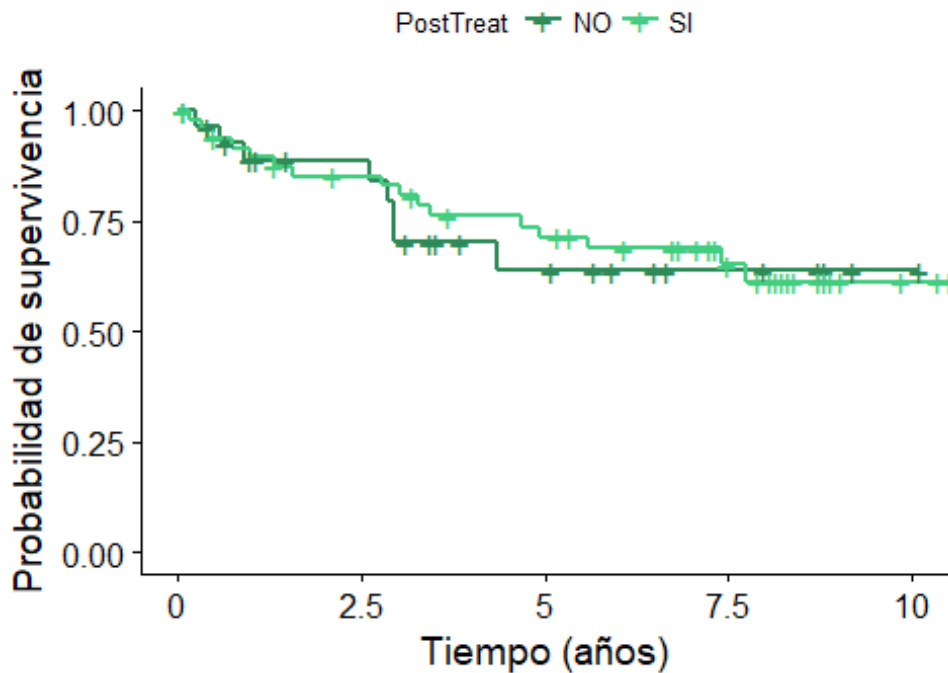
```

```

ggsurvplot(km5, data = Dataset_HCC, title = "Curvas de Supervivencia en
función si se recibió tratamiento tras la cirugía", xlab="Tiempo (años)
", ylab="Probabilidad de supervivencia", main="KM Plot byPostTreat" ,
conf.int = F, legend.labs = c("NO", "SI"), legend.title = "PostTreat", pa
lette = c("seagreen", "seagreen3"))

```

## Curvas de Supervivencia en función si se re



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ PostTreat, data = Dataset_HCC )
```

```
## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
## ~
## PostTreat, data = Dataset_HCC)
##
## n=78, 5 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## PostTreat=NO 29         8     7.46    0.0395    0.0581
## PostTreat=SI 49        16    16.54    0.0178    0.0581
##
## Chisq= 0.1 on 1 degrees of freedom, p= 0.8
```

### #Etiología

```
km6=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Etanol, data = Dataset_HCC )
summary(km6)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
## Etanol, data = Dataset_HCC)
##
## 16 observations deleted due to missingness
##           Etanol=NO
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.247   37     1    0.973  0.0267    0.922    1.000
## 0.411   36     1    0.946  0.0372    0.876    1.000
## 0.575   33     1    0.917  0.0458    0.832    1.000
```

```

## 0.740 31 1 0.888 0.0530 0.790 0.998
## 0.904 30 1 0.858 0.0589 0.750 0.982
## 2.959 26 1 0.825 0.0653 0.707 0.963
## 3.041 25 1 0.792 0.0705 0.665 0.943
## 3.288 23 1 0.758 0.0754 0.623 0.921
## 3.452 22 1 0.723 0.0794 0.583 0.897
## 4.932 18 1 0.683 0.0846 0.536 0.871
##
## Etanol=SI
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164 28 1 0.964 0.0351 0.898 1.000
## 0.329 27 1 0.929 0.0487 0.838 1.000
## 0.986 26 1 0.893 0.0585 0.785 1.000
## 1.315 24 1 0.856 0.0668 0.734 0.997
## 1.562 23 1 0.818 0.0735 0.686 0.976
## 2.795 22 1 0.781 0.0790 0.641 0.953
## 4.685 20 1 0.742 0.0842 0.594 0.927
## 5.589 17 1 0.699 0.0899 0.543 0.899
## 7.397 11 1 0.635 0.1017 0.464 0.869
## 7.726 9 1 0.564 0.1122 0.382 0.833

```

```
print(km6)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muer
te) ~
## Etanol, data = Dataset_HCC)
##
## 16 observations deleted due to missingness
## n events median 0.95LCL 0.95UCL
## Etanol=NO 38 10 NA NA NA
## Etanol=SI 29 10 NA 7.4 NA

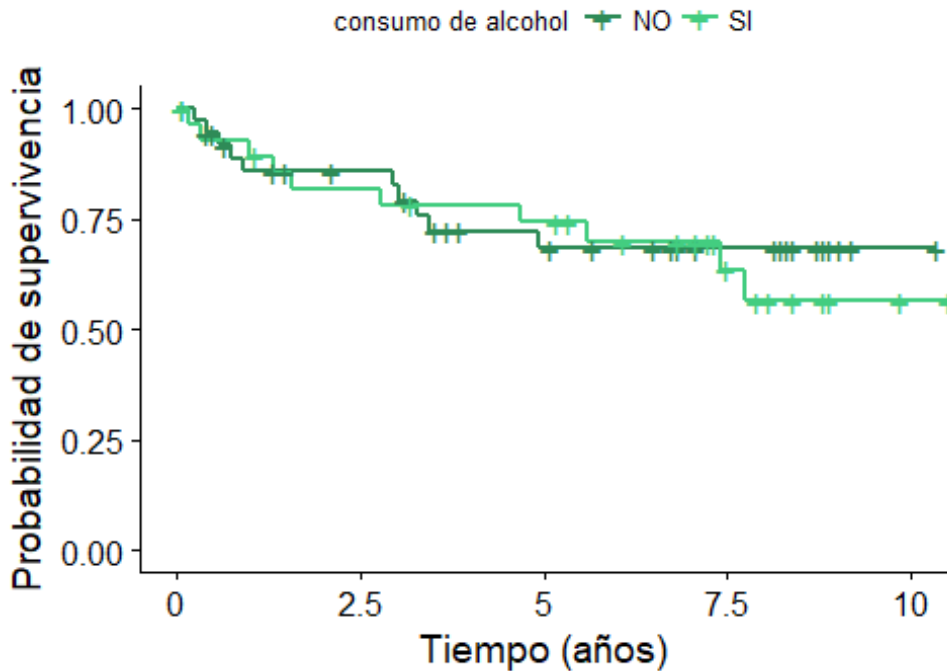
```

```

ggsurvplot(km6, data = Dataset_HCC, title = "Curvas de Supervivencia en
función del consumo de alcohol", xlab="Tiempo (años)", ylab="Probabilid
ad de supervivencia", main="KM Plot by Etanol", conf.int = F, legend.l
abs = c("NO", "SI"), legend.title = "consumo de alcohol", palette = c("s
eagreen", "seagreen3"))

```

## Curvas de Supervivencia en función del con



```
survdif ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Etanol
, data = Dataset_HCC )
```

```
## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
~
## Etanol, data = Dataset_HCC)
##
## n=67, 16 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Etanol=NO 38         10    10.74   0.0516   0.112
## Etanol=SI 29         10     9.26   0.0599   0.112
##
## Chisq= 0.1 on 1 degrees of freedom, p= 0.7
```

```
km7=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Virolog
ia , data = Dataset_HCC )
summary(km7)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
## Virologia, data = Dataset_HCC)
##
## 16 observations deleted due to missingness
##           Virologia=NO
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164    20     1    0.950  0.0487   0.859   1.000
## 0.329    19     1    0.900  0.0671   0.778   1.000
## 0.986    18     1    0.850  0.0798   0.707   1.000
## 1.315    16     1    0.797  0.0908   0.637   0.996
```

```

## 1.562    15     1    0.744  0.0991    0.573    0.966
## 4.685    14     1    0.691  0.1053    0.512    0.931
## 5.589    11     1    0.628  0.1129    0.441    0.893
## 7.397     7     1    0.538  0.1275    0.338    0.856
##
##
##           Virologia=SI
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  0.247   45     1    0.978  0.0220    0.936    1.000
##  0.411   44     1    0.956  0.0307    0.897    1.000
##  0.575   41     1    0.932  0.0378    0.861    1.000
##  0.740   39     1    0.908  0.0437    0.827    0.998
##  0.904   38     1    0.884  0.0487    0.794    0.985
##  2.795   34     1    0.858  0.0538    0.759    0.971
##  2.959   33     1    0.832  0.0581    0.726    0.954
##  3.041   32     1    0.806  0.0618    0.694    0.937
##  3.288   29     1    0.779  0.0656    0.660    0.918
##  3.452   28     1    0.751  0.0689    0.627    0.899
##  4.932   24     1    0.720  0.0728    0.590    0.877
##  7.726   15     1    0.672  0.0823    0.528    0.854

```

```
print(km7)
```

```

## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muer
##   te) ~
##     Virologia, data = Dataset_HCC)
##
## 16 observations deleted due to missingness
##           n events median 0.95LCL 0.95UCL
## Virologia=NO 20     8    NA     5.59    NA
## Virologia=SI 47    12    NA     NA     NA

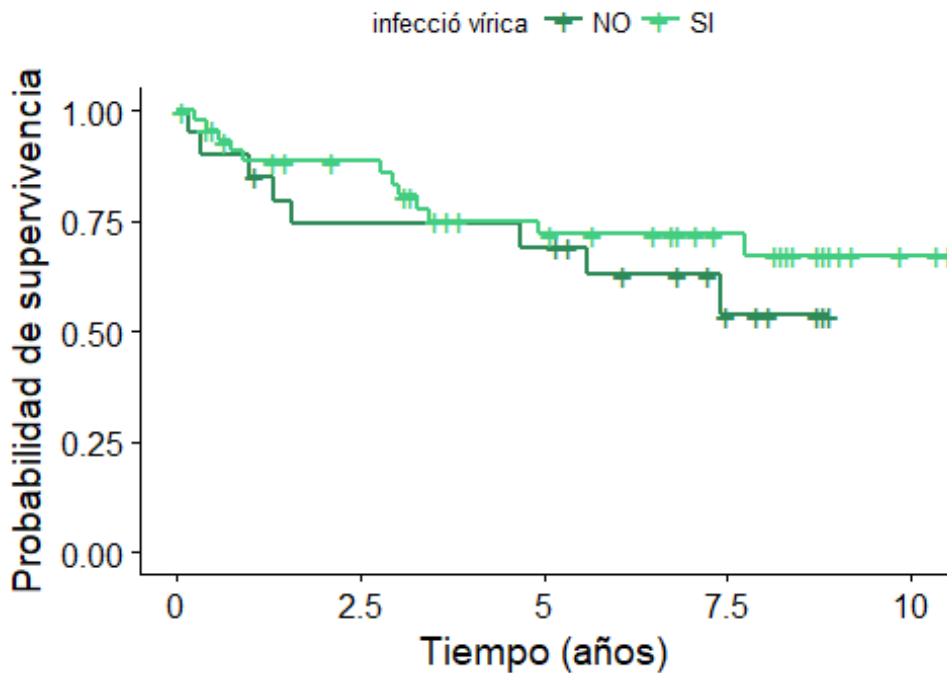
```

```

ggsurvplot(km7, data = Dataset_HCC, title = "Curvas de Supervivencia en
función de si presentaban infección vírica", xlab="Tiempo (años)", ylab
="Probabilidad de supervivencia", main="KM Plot by virologia", conf.int
= F, legend.labs = c("NO", "SI"), legend.title = "infección vírica", pa
lette = c("seagreen", "seagreen3"))

```

## Curvas de Supervivencia en función de si pi



```
survdiff ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Virolog
ia , data = Dataset_HCC )
```

```
## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
~
##   Virologia, data = Dataset_HCC)
##
## n=67, 16 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Virologia=NO 20         8     6.24    0.495    0.722
## Virologia=SI 47        12    13.76    0.225    0.722
##
## Chisq= 0.7  on 1 degrees of freedom, p= 0.4
```

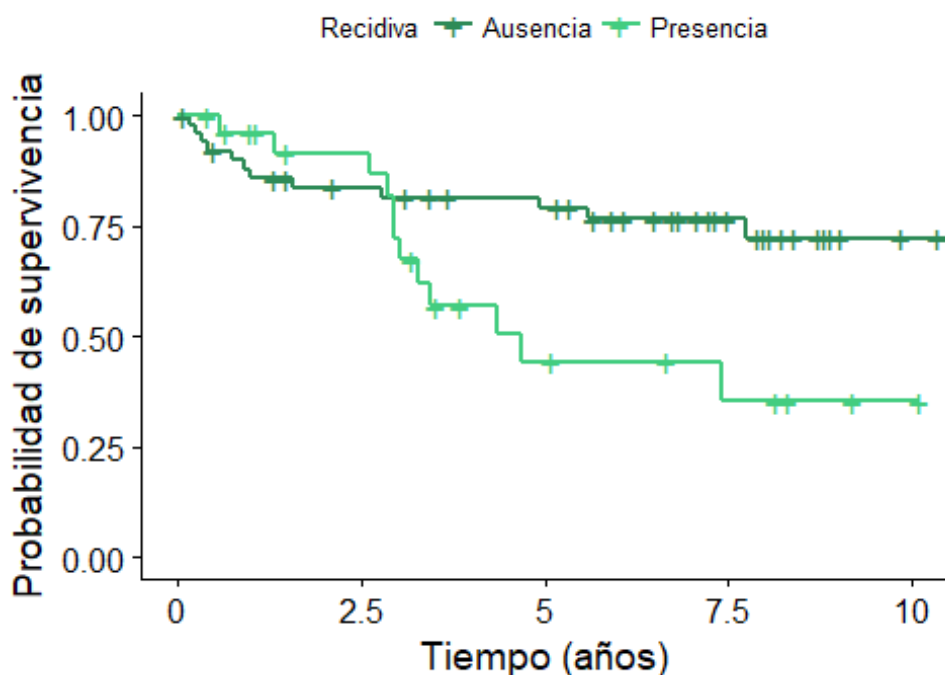
```
##Recidiva
km8=survfit(Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Recidiv
a , data = Dataset_HCC )
summary(km8)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Mu
erte) ~
##   Recidiva, data = Dataset_HCC)
##
## 5 observations deleted due to missingness
##           Recidiva=NO
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.164   50     1    0.980  0.0198   0.942   1.000
## 0.247   49     1    0.960  0.0277   0.907   1.000
## 0.329   48     1    0.940  0.0336   0.876   1.000
```

```
## 0.411 47 1 0.920 0.0384 0.848 0.998
## 0.740 45 1 0.900 0.0426 0.820 0.987
## 0.904 44 1 0.879 0.0463 0.793 0.975
## 0.986 43 1 0.859 0.0495 0.767 0.961
## 1.562 40 1 0.837 0.0527 0.740 0.947
## 2.795 38 1 0.815 0.0558 0.713 0.932
## 4.932 34 1 0.791 0.0590 0.684 0.916
## 5.589 31 1 0.766 0.0624 0.653 0.898
## 7.726 18 1 0.723 0.0720 0.595 0.879
##
##
## Recidiva=SI
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.575 25 1 0.960 0.0392 0.886 1.000
## 1.315 21 1 0.914 0.0582 0.807 1.000
## 2.630 19 1 0.866 0.0723 0.735 1.000
## 2.877 18 1 0.818 0.0828 0.671 0.997
## 2.959 17 2 0.722 0.0971 0.555 0.939
## 3.041 15 1 0.674 0.1018 0.501 0.906
## 3.288 13 1 0.622 0.1064 0.445 0.870
## 3.452 12 1 0.570 0.1094 0.391 0.830
## 4.356 9 1 0.507 0.1141 0.326 0.788
## 4.685 8 1 0.443 0.1161 0.265 0.741
## 7.397 5 1 0.355 0.1221 0.181 0.697
```

```
ggsurvplot(km8, data = Dataset_HCC, title = "Curvas de Supervivencia en
función de la recidiva", xlab="Tiempo (años)", ylab="Probabilidad de su
pervivencia", main="KM Plot by Recidiva", conf.int = F, legend.labs =
c("Ausencia", "Presencia"), legend.title = "Recidiva", palette = c("sea
green", "seagreen3"))
```

## Curvas de Supervivencia en función de la re



```
print(km8)
```

```
## Call: survfit(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~
##   Recidiva, data = Dataset_HCC)
##
##   5 observations deleted due to missingness
##           n events median 0.95LCL 0.95UCL
## Recidiva=NO 52      12      NA      NA      NA
## Recidiva=SI 26      12      4.68    3.29    NA

survdiff ( Surv (Dataset_HCC$Longevidad, Dataset_HCC$Muerte) ~ Recidiv
a , data = Dataset_HCC )

## Call:
## survdiff(formula = Surv(Dataset_HCC$Longevidad, Dataset_HCC$Muerte)
~
##   Recidiva, data = Dataset_HCC)
##
## n=78, 5 observations deleted due to missingness.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Recidiva=NO 52      12      17.02      1.48      5.22
## Recidiva=SI 26      12      6.98      3.62      5.22
##
## Chisq= 5.2 on 1 degrees of freedom, p= 0.02
```

## Modelo de Riesgo proporcionales de Cox.

```
library(flexsurv)
library(survival)
#univariante
coxph(Surv(Longevidad, Muerte) ~ Sexo, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ Sexo, data = Dataset_HCC
)
##
##           coef exp(coef) se(coef)      z      p
## SexoMujer 0.364      1.439      0.472 0.77 0.44
##
## Likelihood ratio test=0.56 on 1 df, p=0.5
## n= 78, number of events= 24
##   (5 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ Edad, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ Edad, data = Dataset_HCC
)
##
##           coef exp(coef) se(coef)      z      p
## Edad>65 0.992      2.697      0.803 1.24 0.22
## Edad50-65 0.584      1.792      0.753 0.78 0.44
##
## Likelihood ratio test=1.83 on 2 df, p=0.4
```



```

## n= 78, number of events= 24
## (5 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ Etanol, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ Etanol, data = Dataset_H
CC)
##
##           coef exp(coef) se(coef)      z      p
## EtanolSI 0.150      1.162    0.448 0.33 0.74
##
## Likelihood ratio test=0.11 on 1 df, p=0.7
## n= 67, number of events= 20
## (16 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ Virologia, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ Virologia, data = Datas
t_HCC)
##
##           coef exp(coef) se(coef)      z      p
## VirologiaSI -0.386      0.680    0.457 -0.84 0.4
##
## Likelihood ratio test=0.69 on 1 df, p=0.4
## n= 67, number of events= 20
## (16 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ EstadioT, data = Dataset_HCC)

## Warning in fitter(X, Y, strats, offset, init, control, weights = we
ights, :
## Ran out of iterations and did not converge

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ EstadioT, data = Datas
t_HCC)
##
##           coef exp(coef) se(coef)      z      p
## EstadioTII 6.36e-01 1.89e+00 6.41e-01 0.99 0.3212
## EstadioTIIIA 1.17e+00 3.22e+00 7.13e-01 1.64 0.1014
## EstadioTIIIB -1.23e+01 4.44e-06 1.81e+03 -0.01 0.9946
## EstadioTIVA 4.85e+00 1.28e+02 1.52e+00 3.19 0.0014
##
## Likelihood ratio test=8.79 on 4 df, p=0.07
## n= 68, number of events= 23
## (15 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ PreTreat, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ PreTreat, data = Datas
t_HCC)
##
##           coef exp(coef) se(coef)      z      p

```

```

## PreTreatSI 0.190      1.209      0.422 0.45 0.65
##
## Likelihood ratio test=0.2 on 1 df, p=0.7
## n= 78, number of events= 24
## (5 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ PostTreat, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ PostTreat, data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## PostTreatSI -0.107      0.898      0.436 -0.25 0.81
##
## Likelihood ratio test=0.06 on 1 df, p=0.8
## n= 78, number of events= 24
## (5 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ HCC, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC, data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## HCC1 y 2  1.679      5.362      0.913 1.84 0.066
## HCC2      0.446      1.562      0.791 0.56 0.573
## HCC2 y 3  2.561     12.952      1.018 2.52 0.012
## HCC3      2.015      7.499      0.800 2.52 0.012
##
## Likelihood ratio test=15.43 on 4 df, p=0.004
## n= 75, number of events= 23
## (8 observations deleted due to missingness)

coxph(Surv(Longevidad, Muerte) ~ Cirrosis, data = Dataset_HCC)

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ Cirrosis, data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## CirrosisSI -0.553      0.575      0.553 -1 0.32
##
## Likelihood ratio test=0.88 on 1 df, p=0.3
## n= 75, number of events= 23
## (8 observations deleted due to missingness)

#Multivariante
coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Sexo+Etanol+Virologia+EstadíoT+Recidiva+PreTreat+PostTreat, data =Dataset_HCC)

## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :
## Ran out of iterations and did not converge

```

```
## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Sexo +
##      Etanol + Virologia + EstadioT + Recidiva + PreTreat + PostTreat
,
##      data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## HCC1 y 2      2.0346   7.6494  1.1399  1.78 0.0743
## HCC2          0.4290   1.5358  0.9480  0.45 0.6509
## HCC2 y 3      3.1127  22.4826  1.3658  2.28 0.0227
## HCC3          1.6235   5.0708  1.0972  1.48 0.1390
## Edad>65       1.8322   6.2477  1.0735  1.71 0.0879
## Edad50-65     1.0811   2.9480  0.9671  1.12 0.2636
## SexoMujer     0.4677   1.5963  0.8474  0.55 0.5810
## EtanolSI      -0.0405   0.9603  0.8634 -0.05 0.9626
## VirologiaSI   -1.0802   0.3395  0.8707 -1.24 0.2148
## EstadioTII     0.7979   2.2209  0.8709  0.92 0.3596
## EstadioTIIIA  1.0503   2.8584  1.1591  0.91 0.3649
## EstadioTIIIB  0.0000   1.0000  0.0000    NA    NA
## EstadioTIVA   6.9171 1009.4334  2.4918  2.78 0.0055
## RecidivaSI    0.7552   2.1280  0.6599  1.14 0.2525
## PreTreatSI    1.1726   3.2305  0.7790  1.51 0.1322
## PostTreatSI   0.0926   1.0970  1.0214  0.09 0.9278
##
## Likelihood ratio test=25.73 on 16 df, p=0.06
## n= 59, number of events= 19
## (24 observations deleted due to missingness)
```

**coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Sexo+Etanol+Virologia+Recidiva+PreTreat+PostTreat, data =Dataset\_HCC)**

```
## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Sexo +
##      Etanol + Virologia + Recidiva + PreTreat + PostTreat, data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## HCC1 y 2      2.438   11.455  1.072  2.27 0.0229
## HCC2          0.791    2.207  0.888  0.89 0.3730
## HCC2 y 3      3.966   52.792  1.332  2.98 0.0029
## HCC3          2.470   11.819  0.968  2.55 0.0107
## Edad>65       1.481    4.395  1.067  1.39 0.1652
## Edad50-65     1.015    2.758  0.956  1.06 0.2884
## SexoMujer     1.325    3.761  0.799  1.66 0.0972
## EtanolSI      0.925    2.523  0.817  1.13 0.2574
## VirologiaSI   -0.548    0.578  0.738 -0.74 0.4581
## RecidivaSI    0.759    2.135  0.600  1.26 0.2064
## PreTreatSI    1.254    3.505  0.628  2.00 0.0457
## PostTreatSI  -0.466    0.628  0.901 -0.52 0.6051
##
## Likelihood ratio test=21.59 on 12 df, p=0.04
## n= 65, number of events= 19
## (18 observations deleted due to missingness)
```

```
coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Sexo+Etanol+Virologia+Recidiva+PreTreat, data =Dataset_HCC)
```

```
## Call:
```

```
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Sexo +  
##      Etanol + Virologia + Recidiva + PreTreat, data = Dataset_HCC)  
##
```

	coef	exp(coef)	se(coef)	z	p
## HCC1 y 2	2.290	9.877	1.021	2.24	0.0249
## HCC2	0.677	1.968	0.859	0.79	0.4307
## HCC2 y 3	3.741	42.161	1.242	3.01	0.0026
## HCC3	2.390	10.916	0.950	2.52	0.0119
## Edad>65	1.619	5.047	1.018	1.59	0.1119
## Edad50-65	0.922	2.515	0.925	1.00	0.3186
## SexoMujer	1.227	3.412	0.757	1.62	0.1049
## EtanolSI	0.773	2.167	0.749	1.03	0.3020
## VirologiaSI	-0.597	0.551	0.732	-0.82	0.4149
## RecidivaSI	0.732	2.079	0.593	1.24	0.2168
## PreTreatSI	1.118	3.058	0.566	1.97	0.0484

```
##
```

```
## Likelihood ratio test=21.33 on 11 df, p=0.03
```

```
## n= 65, number of events= 19
```

```
## (18 observations deleted due to missingness)
```

```
coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Sexo+Etanol+Recidiva+PreTreat, data =Dataset_HCC)
```

```
## Call:
```

```
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Sexo +  
##      Etanol + Recidiva + PreTreat, data = Dataset_HCC)
```

```
##
```

	coef	exp(coef)	se(coef)	z	p
## HCC1 y 2	2.378	10.779	1.040	2.29	0.0223
## HCC2	0.805	2.236	0.852	0.94	0.3449
## HCC2 y 3	3.999	54.561	1.236	3.24	0.0012
## HCC3	2.342	10.405	0.953	2.46	0.0140
## Edad>65	1.629	5.099	1.038	1.57	0.1166
## Edad50-65	1.234	3.434	0.856	1.44	0.1493
## SexoMujer	1.377	3.964	0.738	1.87	0.0621
## EtanolSI	1.066	2.903	0.636	1.67	0.0940
## RecidivaSI	0.822	2.274	0.586	1.40	0.1606
## PreTreatSI	1.112	3.041	0.579	1.92	0.0550

```
##
```

```
## Likelihood ratio test=20.64 on 10 df, p=0.02
```

```
## n= 65, number of events= 19
```

```
## (18 observations deleted due to missingness)
```

```
coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Sexo+Recidiva+PreTreat, data =Dataset_HCC)
```

```
## Call:
```

```
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Sexo +  
##      Recidiva + PreTreat, data = Dataset_HCC)
```

```
##
```

	coef	exp(coef)	se(coef)	z	p
--	------	-----------	----------	---	---

```

## HCC1 y 2      2.341      10.387      1.008  2.32  0.02024
## HCC2          0.775       2.170       0.823  0.94  0.34620
## HCC2 y 3     4.121      61.592      1.223  3.37  0.00075
## HCC3         2.359      10.585       0.842  2.80  0.00509
## Edad>65      1.817       6.151       1.004  1.81  0.07050
## Edad50-65    1.484       4.411       0.877  1.69  0.09076
## SexoMujer    0.719       2.053       0.559  1.29  0.19797
## RecidivaSI   0.772       2.164       0.540  1.43  0.15263
## PreTreatSI   0.958       2.607       0.509  1.88  0.05991
##
## Likelihood ratio test=25.39 on 9 df, p=0.003
## n= 75, number of events= 23
## (8 observations deleted due to missingness)

```

```

coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+Recidiva+PreTreat, data =Dataset_HCC)

```

```

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + Recidiva +
##       PreTreat, data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## HCC1 y 2      2.127      8.392    0.971  2.19  0.0285
## HCC2          0.694       2.001    0.830  0.84  0.4031
## HCC2 y 3     3.878     48.328    1.188  3.26  0.0011
## HCC3         2.168       8.741    0.817  2.65  0.0080
## Edad>65      1.914       6.783    0.990  1.93  0.0531
## Edad50-65    1.475       4.372    0.871  1.69  0.0905
## RecidivaSI   0.708       2.030    0.549  1.29  0.1974
## PreTreatSI   0.954       2.595    0.510  1.87  0.0617
##
## Likelihood ratio test=23.87 on 8 df, p=0.002
## n= 75, number of events= 23
## (8 observations deleted due to missingness)

```

```

coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+PreTreat, data =Dataset_HCC)

```

```

## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + PreTreat,
##       data = Dataset_HCC)
##
##              coef exp(coef) se(coef)      z      p
## HCC1 y 2      2.327     10.245    0.961  2.42  0.01551
## HCC2          0.455       1.576    0.797  0.57  0.56850
## HCC2 y 3     3.717     41.151    1.123  3.31  0.00094
## HCC3         2.148       8.569    0.804  2.67  0.00758
## Edad>65      2.098       8.147    0.955  2.20  0.02802
## Edad50-65    1.390       4.015    0.834  1.67  0.09570
## PreTreatSI   0.728       2.071    0.470  1.55  0.12162
##
## Likelihood ratio test=22.21 on 7 df, p=0.002
## n= 75, number of events= 23
## (8 observations deleted due to missingness)

```

```
CHC.cox<-coxph(Surv(Longevidad, Muerte) ~ HCC+Edad+PreTreat, data =Dataset_HCC)
summary(CHC.cox)
```

```
## Call:
## coxph(formula = Surv(Longevidad, Muerte) ~ HCC + Edad + PreTreat,
##       data = Dataset_HCC)
##
## n= 75, number of events= 23
## (8 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## HCC1 y 2      2.3268   10.2447  0.9614  2.420 0.015511 *
## HCC2          0.4546    1.5756  0.7972  0.570 0.568504
## HCC2 y 3      3.7172   41.1510  1.1233  3.309 0.000936 ***
## HCC3          2.1482    8.5694  0.8045  2.670 0.007579 **
## Edad>65       2.0976    8.1468  0.9548  2.197 0.028021 *
## Edad50-65     1.3900    4.0148  0.8343  1.666 0.095705 .
## PreTreatSI    0.7281    2.0711  0.4703  1.548 0.121618
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## HCC1 y 2          10.245   0.09761   1.5566   67.426
## HCC2              1.576   0.63469   0.3303   7.517
## HCC2 y 3          41.151   0.02430   4.5520  372.016
## HCC3              8.569   0.11669   1.7708  41.470
## Edad>65           8.147   0.12275   1.2540  52.929
## Edad50-65         4.015   0.24908   0.7825  20.598
## PreTreatSI        2.071   0.48283   0.8239   5.207
##
## Concordance= 0.768 (se = 0.063 )
## Rsquare= 0.256 (max possible= 0.91 )
## Likelihood ratio test= 22.21 on 7 df, p=0.002
## Wald test              = 21.17 on 7 df, p=0.004
## Score (logrank) test = 25.44 on 7 df, p=6e-04
```

```
cox.zph(CHC.cox)
```

```
##              rho   chisq    p
## HCC1 y 2     -0.2115  1.03757 0.308
## HCC2         -0.0750  0.12587 0.723
## HCC2 y 3      0.1048  0.25114 0.616
## HCC3         -0.0301  0.01970 0.888
## Edad>65       0.0127  0.00334 0.954
## Edad50-65    -0.0599  0.06098 0.805
## PreTreatSI   -0.2575  1.32635 0.249
## GLOBAL              NA  4.95633 0.665
```

```
ggcoxdiagnostics(CHC.cox, type = "dfbeta")
```

