

FACULTAD DE CIENCIAS EXPERIMENTALES

UNIVERSIDAD DE ALMERÍA

Curso Académico: 2020-2021

# APLICACIONES DEL ANÁLISIS DE SUPERVIVENCIA EN LA BIOTECNOLOGÍA



UNIVERSIDAD  
DE ALMERÍA



TRABAJO FIN DE GRADO

GRADO EN BIOTECNOLOGÍA

Autor: Nicolás Russo Martínez

Directora: Yolanda del Águila del Águila

Departamento de Matemáticas

Área de Estadística e Investigación Operativa

Mayo, 2021

# Índice

Resumen.....	4
Abstract .....	4
Objetivos .....	5
1. Introducción .....	5
2. Censura.....	6
3. Función de supervivencia y función de riesgo .....	8
3.1. Función de supervivencia .....	8
3.2. Función de riesgo .....	9
4. Tabla de vida.....	11
5. Taxonomía del análisis de supervivencia .....	14
5.1. Modelos no paramétricos .....	14
5.1.1. Estimador de Kaplan-Meier.....	14
5.1.2. Estimador de Nelson-Aalen .....	18
5.1.3. Prueba log-rank .....	18
5.2. Modelos semi-paramétricos .....	21
5.2.1. Modelo de los riesgos proporcionales de Cox .....	21
5.3. Modelos paramétricos .....	25
5.3.1. Modelo de Weibull .....	25
5.3.2. Modelo Exponencial.....	29
6. Aplicaciones.....	32
6.1. Medicina y Salud Humana.....	32
6.2. Vegetal.....	39
6.3. Animal.....	42
6.4. Ambiental .....	46
7. Conclusiones.....	46

Bibliografía..... 47

ANEXO TABLAS ..... 50

## Resumen

El análisis de supervivencia es una técnica estadística que a lo largo de los años ha demostrado su utilidad al aplicarse en diversas áreas, en las que la biotecnología no resulta una excepción. Al contrario de lo que inicialmente se esperaría, el análisis de supervivencia no se centra únicamente en la supervivencia como tal, sino que ha visto ampliados sus horizontes al considerar la muerte como un evento, y no sólo restringirse al final de una vida. Estimar la probabilidad de que suceda un evento, el riesgo en el que se encuentra un grupo poblacional en concreto, comparar entre distintos procedimientos o la esperanza de vida de un individuo son incógnitas que pueden ser resueltas mediante la aplicación de las diversas metodologías con las que cuenta el análisis de supervivencia, como el método de Kaplan-Meier, la tabla de vida, o la regresión de Cox. El análisis de supervivencia ha puesto sus técnicas al servicio de la biotecnología, permitiendo estudiar mejores tratamientos para enfermedades, desarrollar fármacos más eficientes, proteger el medio ambiente y el desarrollo de mejores variedades vegetales, entre muchas otras.

## Abstract

Survival analysis is a statistical technique that over the years has proven its usefulness when applied in various areas, in which biotechnology is not an exception. Contrary to what would initially be expected, survival analysis does not focus solely on survival as such, but has seen its horizons broadened by considering death as an event, and not just restricted to the end of a life. Estimating the probability of an event happening, the risk of a specific population group, comparing between different procedures or the life expectancy of an individual are unknowns that can be resolved by applying the various methodologies with which Survival analysis counts, such as the Kaplan-Meier method, the Life Table, or the Cox regression. Survival analysis has put its techniques at the service of biotechnology, allowing the study of better treatments for diseases, developing more efficient drugs, protecting the environment, and developing better plant varieties, among many other.

## Objetivos

El siguiente Trabajo de Fin de Grado tiene como objetivo el estudio de distintos métodos estadísticos relacionados con el análisis de supervivencia y mostrar distintas aplicaciones en el área de la biotecnología. Para ello, se establecieron los siguientes objetivos específicos:

1. Recabar información sobre el análisis de supervivencia y las dificultades que conlleva la censura en los análisis estadísticos tradicionales.
2. Conocer y comprender los modelos y técnicas del análisis de supervivencia.
3. Revisar sus aplicaciones en los distintos campos que componen la biotecnología.
4. Exponer investigaciones en las que la aplicación del análisis de supervivencia resultó indispensable para el correcto desarrollo de éstas.

## 1. Introducción

El análisis de supervivencia es una rama de la estadística cuyos comienzos se remontan a 1662, a manos de John Graunt, pero, a pesar de que en sus inicios su orientación estaba meramente enfocada en estimar la mortalidad, con el paso de los años ha logrado aplicarse en una amplia variedad de áreas, como la ingeniería, el marketing o la investigación biomédica.

Originalmente, el análisis de supervivencia permitía estimar la probabilidad de sobrevivir, la esperanza de vida o el riesgo de muerte de un individuo poseedor de unos atributos definidos, como la edad o el género. Pero con la introducción del análisis de supervivencia en otras áreas, términos como muerte o supervivencia empezaron a adquirir nuevas connotaciones, entendiéndose a partir de entonces la muerte, o el fallo en el caso de la ingeniería, como la manifestación de un evento de interés, sin que este tenga que implicar de ninguna manera alguna clase de perjuicio para el sujeto de estudio, y, por lo tanto, se entiende como supervivencia la ausencia de dicho evento.

Cuando se habla de supervivencia, se habla de probabilidades. Esto significa que la supervivencia puede ser medida como una probabilidad. La supervivencia implica la pertenencia a un grupo, el cual consiste en un número de individuos, a cada uno de los cuales se le asocia una probabilidad de supervivencia. La pertenencia de cada individuo al grupo puede verse alterada debido a un evento como la muerte o una migración.

Usualmente las técnicas estadísticas habituales no pueden emplearse porque la distribución asociada a la variable tiempo de supervivencia, no suele ser una distribución Normal y la información tiende a estar censurada.

## 2. Censura

Una de las cualidades que caracteriza al análisis de supervivencia, y que conlleva que se requieran técnicas estadísticas especiales para su desarrollo, es la posibilidad de la aparición de tiempos incompletos de supervivencia en los individuos sometidos a estudio, a estos tiempos incompletos se los conoce como censura.

Con objeto de realizar un análisis de supervivencia, se debe registrar el tiempo hasta el evento,  $t_i$ , para los individuos  $i \in \{1, \dots, N\}$  de un grupo, pero, en ocasiones, la recolección completa de los datos puede resultar un desafío. Especialmente en estudios con humanos, donde la recolección de datos puede prolongarse durante años, lo cual puede llevar a que parte de la información se pierda, ya sea porque el evento de interés no se llegó a observar, parte de los sujetos abandonaron el estudio o no mantuvieron continuidad en su seguimiento, por ejemplo. Es esta pérdida de información, una de las principales razones por las que no se puede utilizar un modelo de regresión tradicional para realizar predicciones con este tipo de datos (Prinja *et al.*, 2010).

La censura puede clasificarse de manera general en tres categorías (Leung *et al.*, 1997):

- Censura por la derecha: se produce cuando el tiempo observado es menor al verdadero tiempo de supervivencia. Es el tipo de censura que más habitualmente se encuentra en el análisis de supervivencia, al mismo tiempo de ser el más fácil de afrontar. La censura por la derecha se aplica cuando un individuo que está siendo estudiado desde un tiempo inicial  $t_0$ , hasta un tiempo posterior  $t_c$ , aún no ha padecido el evento de interés en el momento de llegar a  $t_c$ . Esta situación puede darse por diversos motivos, como por ejemplo el hecho de que el estudio finalice antes de que el evento se produzca, o el individuo abandona el estudio durante su desarrollo.
- Censura por la izquierda: se produce cuando el tiempo observado es mayor que el verdadero tiempo de supervivencia, en otras palabras, la censura por la izquierda se produce en aquellos casos en los que se sabe que los objetos de estudio han padecido el evento antes de un tiempo específico, pero dicho evento pudo suceder en cualquier momento antes del tiempo de censura.
- Censura por intervalos: se produce cuando se conoce el intervalo entre dos tiempos en el cual se produjo el evento, pero no el momento exacto en el que se produjo.

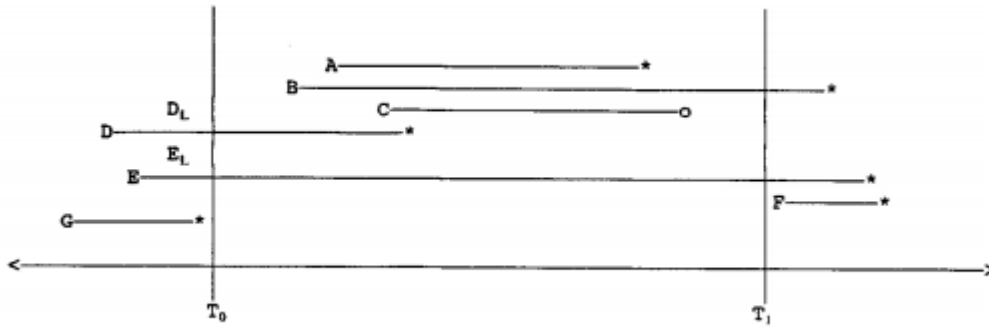


Figura 1: Representación de los distintos tipos de censura donde: A (sin censura), B (censura por la derecha por finalización del estudio), C (censura por la derecha por pérdida del seguimiento), D (censura por la izquierda), E (censura por la derecha y por la izquierda), F y G (completamente censurados por la derecha y la izquierda, respectivamente) (Leung *et al.*, 1997).

A su vez, pueden establecerse otros tres tipos de censura (Lee & Wang, 2003):

- Censura Tipo I: Todos los individuos empiezan y terminan el estudio al mismo tiempo (duración del estudio fijada).
- Censura Tipo II: Todos los sujetos empiezan el estudio al mismo tiempo, pero el estudio termina cuando un número predeterminado de sujetos han experimentado el evento (duración del estudio flexible).
- Censura Tipo III: Los sujetos entran en el estudio en distintos momentos, pero la duración del estudio está fijada.

En presencia de datos con censura, se suelen emplear cuatro métodos estadísticos (Leung *et al.*, 1997): Análisis completo de datos, el enfoque de imputación, análisis basado en variables dicotómicas y el enfoque de similitud.

### Análisis completo de datos

Es un enfoque cuya principal ventaja yace en la simplicidad, puesto que se basa en analizar únicamente aquellas observaciones que se encuentren completas e ignorar las que manifiesten censura. El problema de este enfoque es la pérdida de eficiencia y un sesgo en la estimación.

### Imputación

A pesar de tratarse de una herramienta popular a la hora de tratar con datos incompletos, no resulta tan eficaz con los datos censurados, puesto que para aplicarlo se suele requerir asumir que los datos censurados son fallos o no, lo cual tiende a sobreestimar o subestimar las probabilidades de supervivencia. Para evitar esto, se puede tratar de aplicar un modelo específico al tiempo hasta el evento en aquellos

datos que se encuentren censurados, pero debido a que no se puede comprobar que el modelo elegido sea el adecuado debido a la naturaleza de los datos en los que se aplica, no se puede llegar a saber su veracidad.

### **Análisis basado en variables dicotómicas**

Permite evitar los problemas de la censura por la derecha y la censura por intervalos al analizar la incidencia del evento durante un periodo de tiempo fijado en forma de ocurrencia contra no ocurrencia, permitiendo así analizar fácilmente las variables dicotómicas mediante las técnicas habituales con resultados binarios como las tablas de contingencia. Pero, a pesar de que no permite emplear distintas covariables en el modelado, en aquellos casos en los que el riesgo de fallo sea bajo, los tiempos de supervivencia sean largos y las covariables se relacionen más con prevenir el evento que con prolongar el tiempo de supervivencia, el enfoque mediante variables dicotómicas puede ser aceptable.

### **Enfoque de similitud**

El más efectivo de los enfoques a la hora de hacer frente a los problemas con la censura, ajustando la estimación según si lo observado en un individuo ha sido censurado o no. El empleo de diversos modelos como el modelo de Kaplan-Meier para la estimación permite utilizar toda la información disponible, sin embargo, eso no descarta la necesidad de realizar suposiciones con respecto a la censura.

## **3. Función de supervivencia y función de riesgo**

### ***3.1. Función de supervivencia***

Uno de los objetivos del análisis de supervivencia, consiste en dar información para estimar y representar gráficamente la función de supervivencia.

La función de supervivencia  $S(t)$  determina la proporción de individuos que sobreviven más allá de un periodo determinado de tiempo  $t$ , o lo que es lo mismo, mide la probabilidad de que un individuo sobreviva más allá de un tiempo  $t$ :

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

Siendo  $T \geq 0$ , una variable aleatoria que representa el tiempo de supervivencia (o tiempo hasta el evento de interés).

La función de supervivencia tiene las siguientes propiedades:

- El rango de tiempo es  $t \in [0, \infty)$ .



- $S(t)$  es una función no creciente, eso significa que,  $S(t_1) \geq S(t_2)$  para  $t_1 \leq t_2$
- A un tiempo  $t=0$ ,  $S(t=0)=1$ , por lo que la probabilidad de sobrevivir en un tiempo inicial 0 es 1, lo cual representa el hecho de que al comienzo del estudio todos los individuos están vivos.
- Si  $T$  es una variable aleatoria continua con función de densidad  $f(t)$  y función de distribución  $F(t) = P[T \leq t]$  entonces  $S(t)$  es continua y  $S(t)=1-F(t)$ . Además, puesto que:

$$S(t) = P[T > t] = \int_t^{\infty} f(x)dx ,$$

la función de densidad está relacionada con la función de supervivencia por:

$$f(t) = -\frac{dS(t)}{dt} = -S'(t)$$

El valor esperado del tiempo de supervivencia  $T$ , que representa la esperanza de vida o vida media de los individuos del grupo se obtiene mediante:

$$\mu = E[T] = \int_0^{\infty} tf(t)dt$$

Sabiendo que  $f(t) = -S'(t)$ , se puede realizar una integración por partes, para lograr de esa manera obtener la esperanza de vida empleando la función de supervivencia  $S(t)$ .

$$\mu = E[T] = \int_0^{\infty} S(t)dt$$

### 3.2. Función de riesgo

La función de riesgo  $h(t)$  es la probabilidad condicional de que suceda el evento en el momento  $t$  habiendo sobrevivido hasta dicho momento. La función de riesgo  $h(t)$ , tiene el significado de un riesgo instantaneo, puesto que para  $\Delta t \rightarrow 0$  solo son considerados los individuos con

$$\begin{aligned} T &\geq t \\ T &< t + \Delta t \end{aligned}$$

La función de riesgo es la tasa de variación instantánea a la que los eventos están sucediendo a los individuos que sobreviven a un tiempo  $t$ . Por lo que podemos escribir la función de riesgo como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

La diferencia entre una función de supervivencia y una función de riesgo se puede resumir en:

- La función de riesgo se centra en la probabilidad de que ocurra el evento en un instante de tiempo  $t$
- La función de supervivencia se centra en la probabilidad de supervivencia en un tiempo  $t$ .

La función de riesgo  $h(t)$  puede ser determinada como el cociente entre la función de densidad y la función de supervivencia:

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

La función de riesgo  $h(t)$  tiene las siguientes propiedades

- $h(t) \geq 0$  para todo  $t$
- $h(t)$  no tiene limite superior
- $h(t)$  puede asumir cualquier forma

Y finalmente la función de riesgo acumulado describe el riesgo acumulado hasta un tiempo  $t$  como:

$$H(t) = \int_0^t h(x)dx$$

Relacionandose con la función de supervivencia de la siguiente manera:

$$S(t) = e^{-H(t)}$$

De modo tal que, a mayor riesgo, menor es la supervivencia. (Kleinbaum y Klein., 2010).

Aquellos análisis de supervivencia en los que puede haber más de una causa de fallo o evento, siendo en ocasiones la manifestación de uno de los eventos un impedimento para la aparición de los demás se consideran como riesgos competitivos (Putter *et al.*, 2007).

Para hacer frente a la presencia de riesgos competitivos, el método más aplicado es el modelo de riesgo de causa específica. Siendo  $T_k$  la variable aleatoria que representa el tiempo desde la causa  $k$  hasta el momento del evento, para  $k=1, \dots, K$ , donde  $K$  representa el distinto número de eventos. Debido a las características de los riesgos competitivos, el único tiempo de supervivencia observado  $T = \min \{T_1, T_2, \dots, T_K\}$ , suele corresponder a la primera causa, junto con su indicador  $\delta=k$  cuando el individuo experimenta el evento debido a la causa  $k$ , lo cual impide la observación del resto de eventos posibles.

El modelo de riesgo de causa específica para la causa  $k$  se define como:

$$h_k(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \delta = k | T \geq t)}{\Delta t}$$

La inferencia de cada  $h_k(t)$  considera como censurados al resto de eventos distintos al observado debido al evento  $k$ .

## 4. Tabla de vida

El método de la tabla de vida lleva empleándose desde hace siglos, el cual, además de ser sencillo, permite representar la mortalidad de un grupo de individuos en forma de probabilidades.

Las tablas de vida pueden emplearse para estudiar e incluso predecir el tamaño poblacional, la fertilidad, la reproducción y la migración (Pavia *et al.*, 2012).

Una tabla de vida consiste en una serie de filas y columnas, cada fila representa un periodo discreto de tiempo, como un año, el cual puede a su vez vincularse a otras cualidades como el sexo o la raza; y las columnas proporcionan información respecto a la mortalidad.

A un tiempo  $j$ , y un punto de inicio  $n_0$  que representa los individuos vivos al comienzo, las columnas de una tabla de vida se componen de (Muriel *et al.*, 2010):

- $n_j$ : número de individuos que sobreviven al comienzo de cualquier edad  $j$ , comenzando desde el número inicial  $n_0$
- $d_j$ : número de individuos que mueren habiendo llegado a  $j$ , pero antes de alcanzar  $j+1$

$$n_{j+1} = n_j - d_j$$

- $c_j$ : número de individuos censurados a una edad  $j$
- $q_j$ : probabilidad de que un individuo muera antes de llegar a  $j+1$

$$q_j = \frac{d_j}{n_j}$$

- $p_j$ : probabilidad de que un individuo sobreviva hasta llegar a  $j+1$

$$p_j = 1 - q_j$$

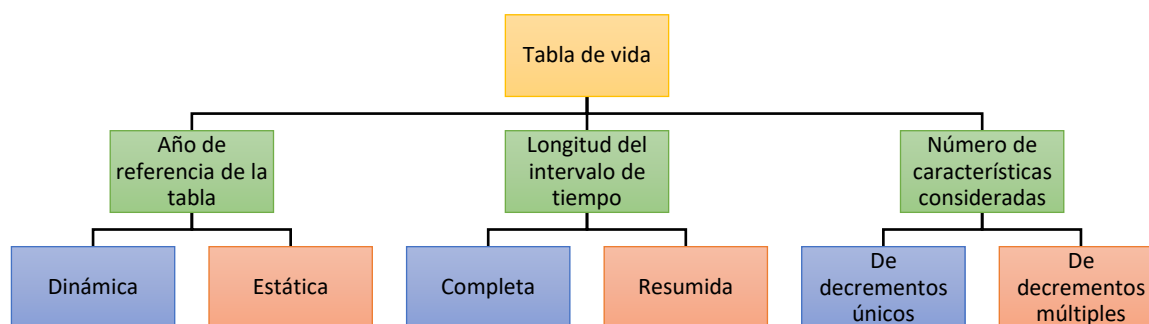
### Ejemplo demostrativo:

Para la realización de este ejemplo demostrativo y todos los posteriores, tomaremos los datos recopilados por Freireich *et al.* (1963), donde estudiaron la efectividad de un nuevo tratamiento en 42 pacientes con leucemia. Originalmente el estudio se realizó a lo largo de 35 semanas, pero para mostrar este ejemplo, emplearemos solo las primeras 10 semanas.

$j$	$n_j$	$d_j$	$c_j$	$q_j$	$p_j$
1	42	2	0	0,048	0,952
2	40	2	0	0,050	0,950
3	38	1	0	0,026	0,974
4	37	2	0	0,054	0,946
5	35	2	0	0,057	0,943
6	33	3	1	0,121	0,879
7	29	1	0	0,034	0,966
8	28	4	0	0,143	0,857
9	24	0	1	0,042	0,958
10	23	1	1	0,087	0,913

Tabla 1: Tabla de vida obtenida de los datos recopilados para estimar la eficacia de un nuevo tratamiento de leucemia en los pacientes.

Las tablas de vida pueden clasificarse según (Etikan *et al.*, 2017):



- **Tabla de vida dinámica**

También llamada horizontal o de cohorte, muestra la información de un grupo desde su nacimiento hasta el momento en el que ninguno de los individuos que lo compone permanece con vida. A pesar de la ventaja de su simplicidad, cuando se emplea con seres vivos con una larga esperanza de vida requiere la observación por un largo periodo de tiempo, lo que establece a la tabla de vida dinámica como más adecuada para seres vivos con una esperanza de vida más breve como los insectos y algunas plantas.

- **Tabla de vida estática**

También conocida como estacional o vertical, representa a una generación entera mediante el empleo de una sección transversal única de tiempo. Su principal ventaja es que proporciona información en un periodo específico de tiempo, pero posee en contrapunto una elevada complejidad conceptual.

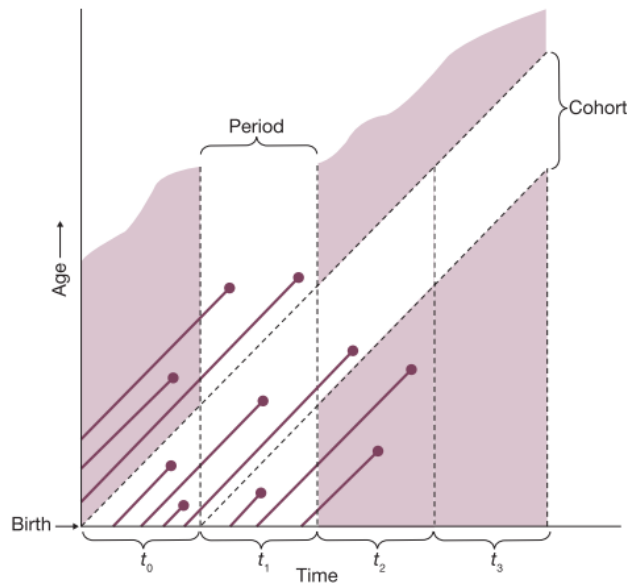


Figura 2: Comparación de la obtención de datos de una tabla de vida estática y una dinámica

- **Tabla de vida completa**

Incluye toda la información de cada año, desde el nacimiento hasta la muerte del individuo.

- **Tabla de vida resumida**

La más usada habitualmente, su información aparece en intervalos de tiempo, a excepción de los primeros años.

- **Tabla de vida de decremento único**

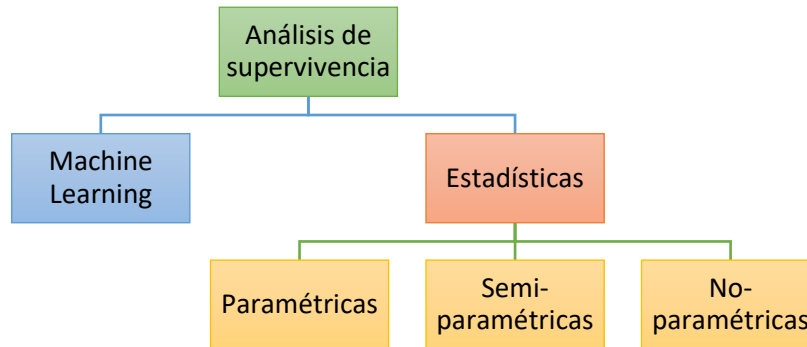
En esta tabla solo se toma en cuenta una característica y una causa de muerte.

- **Tabla de vida de decremento múltiple**

Incluye más de una causa de muerte y más de una característica, permitiendo así representar los efectos separados y combinados de estas.

## 5. Taxonomía del análisis de supervivencia

Las técnicas del análisis de supervivencia pueden dividirse en estadísticas y en Machine Learning, pero a lo largo del desarrollo de esta memoria, nos centraremos en las técnicas estadísticas. Las técnicas estadísticas a su vez se dividen en otras tres categorías: paramétricas, semi-paramétricas y no-paramétricas.



Uno de los objetivos del análisis de datos de supervivencia puede ser examinar si los tiempos de supervivencia están relacionados con otros rasgos. Los modelos de regresión pueden ser empleados para evaluar el efecto de covariables en los resultados.

### 5.1. Modelos no paramétricos

Los métodos no paramétricos se pueden emplear en aquellos casos en los que resulte difícil hacer que los datos se ajusten a una distribución teórica, puesto que no se realizan suposiciones que deban seguirse ni existe una distribución subyacente al evento temporal. Los métodos no paramétricos resultan muy útiles para resumir los datos de supervivencia y realizar comparaciones simples, pero no logran funcionar correctamente en situaciones más complejas.

#### 5.1.1. Estimador de Kaplan-Meier

El estimador de Kaplan-Meier, también conocido como estimador límite-producto, se asemeja a una variación de la tabla de vida, en la cual el eje del tiempo se divide en varios intervalos discretos de tiempo

contabilizando factores como el número de individuos que sobreviven al comienzo de cada intervalo, el número de individuos que muere en cada intervalo y el número de individuos censurados o perdidos en el intervalo. El estimador de Kaplan-Meier también divide el eje del tiempo en varios intervalos discretos. Sin embargo, en este caso los intervalos no se definen por una longitud fija sino por la aparición de un evento. Durante un análisis empleando el estimador de Kaplan-Meier, debe asumirse que la censura es independiente de la probabilidad de riesgo y que se sabe con exactitud en qué momento se produce cada evento (Kleinbaum y Klein, 2010).

El estimador de Kaplan-Meier es un método no paramétrico ampliamente empleado para estimar la función de supervivencia, permitiendo obtener la curva de Kaplan-Meier (Kaplan-Meier, 1958). El método se basa en la idea de que la probabilidad de sobrevivir  $k$  o más periodos desde la entrada al estudio es el producto de  $k$  tasas de supervivencia observadas para cada periodo:

$$S(k) = p_1 \times p_2 \times \dots \times p_k$$

$p_1$  es la proporción de individuos que sobreviven al primer periodo,  $p_2$  es la proporción de individuos que sobreviven más allá del segundo periodo, con la condición de haber sobrevivido hasta el segundo periodo, y así sucesivamente hasta llegar a  $p_k$ .

El estimador de Kaplan-Meier aplica la ley multiplicativa de las probabilidades para calcular la probabilidad de supervivencia en el tiempo  $t_j$  (Massons, 1992). Siendo  $t_1 < t_2 < \dots < t_k$  los tiempos observados del evento, y  $n = n_0$  el tamaño de la muestra. Siendo  $d_j$  el número de individuos que tienen un evento en un tiempo  $t_j$ , donde  $j=1, \dots, k$ , y  $c_j$  el número de individuos censurados en el intervalo  $[t_j, t_{j+1})$ , entonces:

$$n_j = n - [(c_j + d_j) + \dots + (c_k + d_k)]$$

Donde  $n_j$  es el número de individuos en riesgo justo antes de  $t_j$

El estimador de Kaplan-Meier de una función de supervivencia  $S(t)$  se obtiene considerando que la probabilidad condicionada de supervivencia será la complementaria a la probabilidad de riesgo, de modo tal que:

$$\hat{S}(t) = \prod_{j: t_j \leq t} \frac{n_j - d_j}{n_j}$$

Este estimador se mantiene para todos los  $t > 0$  y solo depende de  $n_j$  y  $d_j$  que son:

- $n_j$ : número en riesgo en el tiempo  $t_j$ .
- $d_j$ : número de eventos en el tiempo  $t_j$ .

Aquí  $n_j$  corresponde con el número de sujetos presentes en el tiempo  $t_j$ . Aquellos sujetos que experimenten el evento o se les aplique censura dejan de estar presentes. El problema de este estimador es que solo considera los eventos  $j$  que ocurren antes del tiempo  $t$ , siendo  $t_j < t$ . Por lo tanto, la curva de supervivencia para el tiempo  $t$  considera todos los eventos que sucedieron antes de  $t$  (Akbar *et al.*, 2009).

Es relevante considerar que, cuando se trabaja con un estimador de Kaplan-Meier, los eventos que ocurren en  $\{t_j\}$  son importantes. Esto quiere decir que, entre dos eventos  $t_j$  y  $t_{j+1}$ , la curva de supervivencia es constante. Esto permite reformular el estimador de Kaplan-Meier con una secuencia recursiva como:

$$\hat{S}(t_k) = \frac{n_{k-1} - d_{k-1}}{n_{k-1}} \hat{S}(t_{k-1})$$

Los errores estandar,  $\sqrt{\hat{V}\hat{S}(t)}$ , se pueden calcular mediante la fórmula de Greenwood (Greenwood, 1926), la cual aproxima la varianza como:

$$\hat{V}\hat{S}(t) = \{\hat{S}(t)\}^2 \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

El método de Kaplan-Meier permite también determinar la mediana de supervivencia, la cual, al contrario que la media de supervivencia, no requiere conocer los tiempos de supervivencia completos de todos los individuos. La mediana indica el tiempo que requiere el 50% del conjunto total de la población en estudio en tener el evento, de modo tal que, para obtenerla se estima mediante el percentil 50 de la distribución, a la que corresponde una probabilidad acumulada de supervivencia menor o igual a 0,5.

### Ejemplo demostrativo (Tabla 1 de supervivencia en ANEXO TABLAS)

Continuando con los datos tomados por Freireich *et al.* (1963), mediante el programa estadístico SPSS, analizaremos la supervivencia de los individuos del estudio con el método de Kaplan-Meier.

En primer lugar, hay que establecer las variables con las cuales trabajaremos para la construcción de la tabla, en este caso son: la variable tiempo de supervivencia (medido en semanas), a la que etiquetaremos como TIEMPO; la variable estado, etiquetada como ESTADO, la cual nos indicará mediante un sistema binario (0=censurado, 1=recaída) en que momentos del estudio se produce un evento, y la variable grupo, etiquetada como GRUPO, la cual nos permitirá agrupar a los individuos del estudio según si han recibido el nuevo tratamiento o no (placebo=1, tratamiento=0).

Una vez establecidas las variables haremos: **Analizar** → **Superviv.** → **Kaplan-Meier...** Donde determinaremos la Hora con nuestra variable TIEMPO, el Estado con la variable ESTADO, definiendo el evento con un valor único de 1. En **Opciones**, podemos establecer que estadísticos y gráficos deseamos



obtener, seleccionamos los Estadísticos **Tablas de Supervivencia y Media y mediana de supervivencia** y los gráficos de **Supervivencia y Riesgo**, con lo que obtenemos:

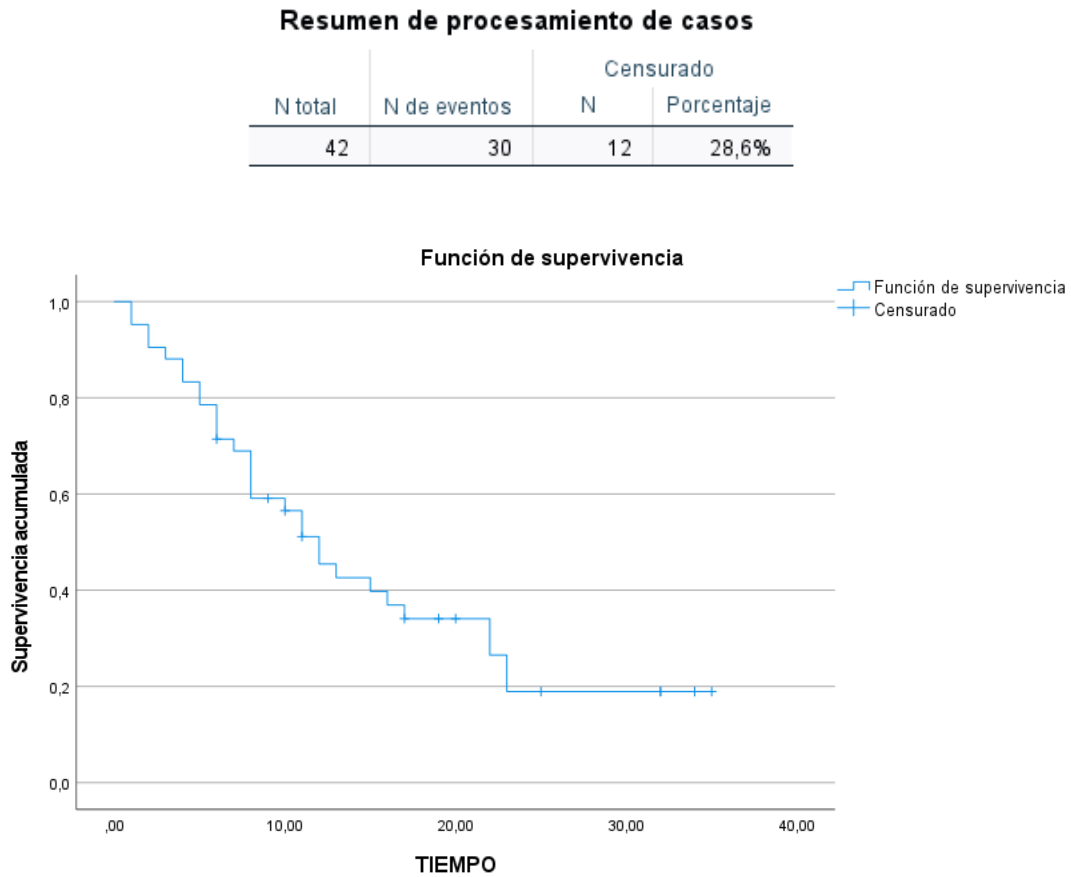


Figura 3: Curva de supervivencia de los pacientes de leucemia (Freireich *et al.*, 1963)

**Medias y medianas para el tiempo de supervivencia**

Estimación	Desv. Error	Media <sup>a</sup>		Estimación	Desv. Error	Mediana	
		Intervalo de confianza de 95 %				Intervalo de confianza de 95 %	
		Límite inferior	Límite superior			Límite inferior	Límite superior
15,339	1,860	11,693	18,985	12,000	1,717	8,636	15,364

a. La estimación está limitada al tiempo de supervivencia más largo, si está censurado.

Figura 4: Estimación de la media y la mediana para el tiempo de supervivencia obtenidas por el método de Kaplan-Meier

### 5.1.2. Estimador de Nelson-Aalen

Al contrario que el estimador de Kaplan-Meier, que es un estimador directo para  $S(t)$ , el estimador de Nelson-Aalen (sugerido por Nelson (1972) y propuesto por Aalen (1978)) es un estimador indirecto para  $S(t)$ . En su lugar, el estimador de Nelson-Aalen realiza una estimación directa de la función de riesgo acumulado dada por:

$$\hat{H}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j} = \sum_{j:t_j \leq t} \hat{h}_j$$

Si  $j=1, \dots, k$ , entonces:

- $d_j$ : número de individuos que tienen un evento en un tiempo  $t_j$ .
- $n_j$ : número de individuos en riesgo justo antes de  $t_j$ .

De aquí se obtiene una estimación para  $S(t)$  mediante

$$\hat{S}(t) = e^{(-\hat{H}(t))} = e^{\left(-\sum_{j:t_j \leq t} \frac{d_j}{n_j}\right)}$$

La estimación de la varianza de Nelson-Aalen es:

$$\hat{\sigma}_{\hat{H}}^2(t) = \sum_{t_j \leq t} \frac{d_j}{n_j^2}$$

### 5.1.3. Prueba log-rank

La prueba log-rank (Le, 1997), también conocida como la prueba Mantel-Haenzel, se puede emplear para determinar si las curvas de supervivencia de los individuos de dos o más grupos distintos, difieren de manera significativa, y, si se aproxima a la distribución chi cuadrado  $\chi^2$ , comprobando si las hipotéticas funciones de supervivencia  $S_0(t), \dots, S_p(t)$  son iguales, basándose en  $(p+1)$  muestras de cada población. Si  $h_j$  denota el riesgo (la probabilidad condicional de fallo) en un tiempo  $t_j$ , la hipótesis nula asociada con la prueba log-rank es que  $h_j$  es común para todas las muestras  $p+1$ . La prueba log-rank compara el número de fallos observados con los esperados y tiene una distribución asintótica  $\chi^2$  bajo la hipótesis nula. Con un número de grupos  $p$ , los grados de libertad son  $p-1$ .

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \dots + \frac{(O_p - E_p)^2}{E_p}$$

Donde:

- $O_p$ : número total de eventos observados en los grupos
- $E_p$ : número total de eventos esperados.

El número total de eventos esperados de un grupo es la suma del número de eventos esperados al momento de cada evento. El número esperado de eventos en el momento de un evento puede calcularse como el riesgo de fallo en ese momento multiplicado por el número de individuos vivos en el grupo.

La prueba log-rank es una prueba de hipótesis no-paramétrica, que en el caso de dos grupos permite resolver el contraste:

$$\begin{cases} H_0: S_1(t) = S_2(t) \\ H_1: S_1(t) \neq S_2(t) \end{cases} \quad \forall t$$

Para ello, aplica los siguientes supuestos:

- Tanto los sujetos censurados como los no censurados tienen la misma probabilidad de padecer el evento, por lo que la censura no es informativa.
- Las curvas Kaplan-Meier de los grupos no deben tener intersecciones entre sí (debe mantenerse la suposición de los riesgos proporcionales).
- Se asume que la curva tiene una distribución libre, por lo que no se le aplica ninguna distribución en particular.

### Ejemplo Demostrativo (Tabla 2 de supervivencia en ANEXO)

En el anterior ejemplo, analizamos los datos de Freireich *et al.* (1963) como un único conjunto, pero, en esta ocasión, analizaremos los datos agrupándolos según si recibieron el nuevo tratamiento o el placebo, para comprobar si existen diferencias en los resultados obtenidos y cómo afectan a la supervivencia de cada grupo.

Los primeros pasos son prácticamente los mismos que en el ejemplo anterior: **Analizar** → **Superviv.** → **Kaplan-Meier**.... Donde establecimos la Hora con nuestra variable TIEMPO, el Estado con la variable ESTADO con un valor único de 1, pero, en esta ocasión añadiremos además la variable GRUPO como Factor. En **Comparar factor**... seleccionamos el estadístico de prueba **Log rango**, y en **Opciones** seleccionamos los Estadísticos **Tablas de Supervivencia** y **Media y mediana de supervivencia**, y los gráficos de **Supervivencia** y **Riesgo**, con lo que obtenemos:

### Resumen de procesamiento de casos

GRUPO	N total	N de eventos	Censurado	
			N	Porcentaje
TRATAMIENTO	21	9	12	57,1%
PLACEBO	21	21	0	0,0%
Global	42	30	12	28,6%

### Comparaciones globales

	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	16,793	1	<,001

Figura 5: Prueba de igualdad de funciones de supervivencia para los distintos niveles de GRUPO

Como podemos observar en la Figura 5, el nivel de significación de la prueba log-rank es  $<0,05$ , por lo que la prueba es significativa, rechazando la hipótesis nula, la cual implica que el tipo de tratamiento recibido no influye en los pacientes, de modo tal que la supervivencia se ve afectada según si reciben el nuevo tratamiento o el placebo, y, como puede observarse en la Figura 6, aquellos pacientes que recibieron el nuevo tratamiento, vieron sus probabilidades de supervivencia incrementadas con respecto a aquellos que sólo recibieron el placebo.

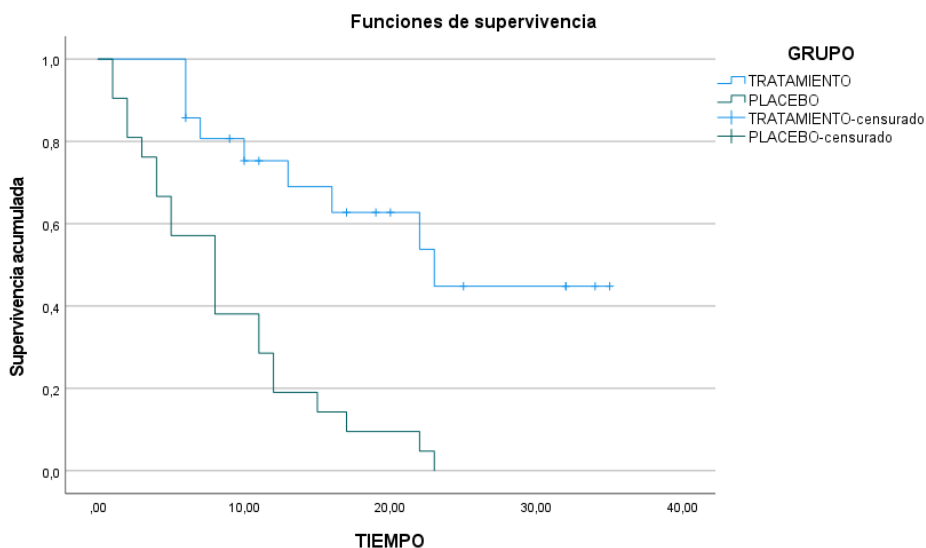


Figura 6: Comparación de las funciones de supervivencia obtenidas por la aplicación de dos tratamientos distintos

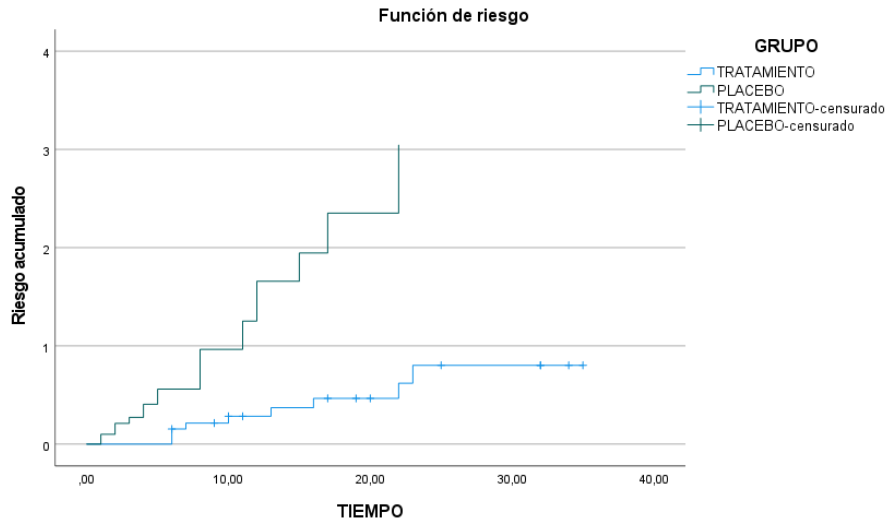


Figura 7: Comparación de las funciones de riesgo obtenidas por la aplicación de dos tratamientos distintos

## 5.2. Modelos semi-paramétricos

Los modelos semi-paramétricos se consideran como un híbrido de los modelos paramétricos y los no-paramétricos. Al igual que los modelos no-paramétricos, los modelos semi-paramétricos no requieren conocer la distribución subyacente del evento en el tiempo.

### 5.2.1. Modelo de los riesgos proporcionales de Cox

Dentro de los modelos semi-paramétricos, el más utilizado es el modelo de los riesgos proporcionales de Cox, el cual asume que los parámetros poseen un efecto multiplicativo en la función del riesgo que es constante sobre el tiempo (Cox, 1972).

El modelo de regresión de los riesgos proporcionales de Cox es un modelo semi-paramétrico porque no realiza suposiciones respecto a la distribución de la probabilidad de riesgo  $h_0(t)$ , sin embargo, asume una forma paramétrica para el efecto de los predictores de riesgo.

El modelo de regresión de Cox es análogo a un modelo de regresión lineal y permite probar la diferencia entre los tiempos de supervivencia de distintos grupos, al mismo tiempo que admite la inclusión de otros factores. En este modelo, la variable dependiente, o de respuesta, es el riesgo (Julia, 2012).

El modelo de riesgos proporcionales de Cox se expresa por:

$$h(t; x) = h_0(t)e^{\beta x}$$

Donde:

- $h_0(t)$ : riesgo base
- $x$ : covariable
- $\beta$ : parámetro a estimar que representa el efecto de la covariable en el resultado

Equivalentemente,

$$\ln h(t; x) = \ln h_0(t) + \beta x$$

El riesgo base es el riesgo en el cual, en caso de haber una sola covariable, la covariable es igual a 0. La principal suposición implicada es la suposición de riesgos proporcionales, que es que la proporción de riesgos (proporción de la función de riesgos para el riesgo base) es constante en el tiempo. El uso de la función exponencial asegura que el riesgo es positivo.

La cantidad de riesgo que se estima a partir del modelo de riesgos proporcionales de Cox es interpretada como un riesgo relativo más que como absoluto. Las covariables se asume que tienen un efecto aditivo sobre el logaritmo natural de la proporción de riesgos. La interpretación del parámetro  $\beta$  es que por cada incremento unitario en la covariable  $x$ , el riesgo se multiplica por  $e^\beta$ .

Como en todos los modelos de regresión, más de una variable puede ser incluida en el modelo de riesgos proporcionales de Cox para lograr así ajustar el efecto de otras variables. La versión multivariable del modelo de riesgos proporcionales de Cox es:

$$h(t; x_1, \dots, x_p) = h_0(t) e^{\beta_1 x_1 + \dots + \beta_p x_p}$$

Equivalentemente,

$$\ln h(t; x_1, \dots, x_p) = \ln h_0(t) + \beta_1 x_1 + \dots + \beta_p x_p$$

En este caso, el riesgo base es el riesgo para un individuo con todas sus covariables iguales a 0 ( $x_1 = \dots = x_p = 0$ ). El efecto de la variable  $x_k$  es interpretado como sigue: por cada incremento unitario en  $x_k$  y manteniendo todas las demás covariables fijas, el riesgo es multiplicado por  $e^{\beta_k}$ .

Los parámetros  $\beta$  en un modelo de riesgos proporcionales de Cox pueden ser estimados mediante la probabilidad parcial tomando el producto sobre eventos temporales ordenados:

$$L(\beta) = \prod_{t_i: \text{evento en } t_i} \frac{h_0(t) e^{\beta x(t_i)}}{\sum_{j \in R(t_i)} h_0(t) e^{\beta x_j}}$$

Donde:

- $R_{(t_i)}$ : Riesgo fijado a un tiempo  $t_i$ .
- $x_{(t_i)}$ : Valor de  $x$  para los sujetos que tuvieron un evento en  $t_i$ .

La probabilidad parcial es la probabilidad de que un individuo que ha tenido un evento en un momento dado es, en realidad, el individuo que lo ha tenido de entre todos los individuos en riesgo en ese momento. Puesto que el riesgo base  $h_0$  está presente tanto en el numerador como en el denominador, se puede simplificar la probabilidad parcial, quedando:

$$L(\beta) = \prod_{t_i: \text{evento en } t_i} \frac{e^{\beta x_{(t_i)}}}{\sum_{j: t_j \geq t_i} e^{\beta x_j}}$$

Es por esto por lo que el modelo de riesgos proporcionales de Cox se establece como un método semi-paramétrico, de modo que los tiempos de supervivencia se asume que están relacionados con variables explicativas de una manera particular, pero no se realizan suposiciones sobre la forma general de los tiempos de supervivencia.

La probabilidad parcial puede ser tratada como una probabilidad. Los errores estándar por la estimación de  $\beta$  se basan en resultados asintóticos. (Kartsonaki, 2016).

La plausibilidad de las suposiciones de los riesgos proporcionales debe comprobarse. Puede comprobarse gráficamente o mediante la inclusión de un efecto dependiente del tiempo y examinando su significación. Aplicando los residuos de Schoenfeld se comprueba si el efecto  $\beta$  como función del tiempo varía con el tiempo. Si varía de manera significativa con el tiempo, implicaría que la suposición de los riesgos proporcionales puede no ser plausible. Existe una prueba  $\chi^2$  derivada de los residuos de Schoenfeld que prueba las desviaciones de la hipótesis nula de proporcionalidad, con un resultado significativo siendo evidencia de no proporcionalidad.

### 5.2.2.1 Variables explicativas dependientes del tiempo

El valor de una variable explicativa, a pesar de ser considerada como variable independiente, puede variar con el tiempo, siendo en ese caso llamada variable explicativa dependiente del tiempo, variable explicativa variante con el tiempo o variable explicativa actualizada con el tiempo. El modelo de los riesgos proporcionales de Cox puede adaptarse a dichas variables. En la práctica esto requeriría separar el tiempo en unidades discretas, como por ejemplo meses, y asignando un valor para cada variable en cada unidad de tiempo durante la cual los individuos están en riesgo. Por lo tanto, la información relacionada con variables explicativas dependientes del tiempo se representa con múltiples anotaciones para cada individuo, cada una representando una unidad de tiempo durante la cual el individuo estaba en riesgo.

Una variable explicativa dependiente del tiempo es diferente de un efecto dependiente del tiempo. Un modelo de riesgos proporcionales de Cox con una variable dependiente del tiempo puede escribirse como:

$$h(t; x(t)) = h_0(t)e^{\beta x(t)}$$

Esto otorga un efecto en el riesgo en un tiempo  $t$  de la variable explicativa en ese tiempo  $t$ . Es posible tener diferentes formulaciones, permitiendo por ejemplo variables retardadas en el tiempo. El efecto en los datos de una variable que se incrementa por una unidad por cada unidad de tiempo, es el mismo que en la versión no actualizada de la variable, dado que su efecto en el logaritmo natural de la proporción de riesgos es lineal.

Para variables que cambian con el tiempo pero solo se miden un número de veces limitado durante el tiempo de seguimiento, es común mantener sus valores más recientes como valores actuales. Sin embargo, debe tenerse precaución en caso de que la variable se mida en un tiempo dado, puesto que puede depender de factores que también puedan influir en su valor y el resultado. (Kartsonaki, 2016).

### 5.2.2.2 Modelos de fragilidad

Los modelos de fragilidad se emplean para incorporar heterogeneidad entre individuos empleando un efecto aleatorio.

Una de las mayores suposiciones del modelo de Cox es la independencia de los tiempos de supervivencia de los distintos individuos. Lo que no impide que algunos factores como el hecho de formar parte de un mismo entorno generen cierta dependencia entre los tiempos de supervivencia. Si se lleva a cabo un análisis de supervivencia ignorando estos factores, el resultado obtenido se verá alterado, afectando a su fiabilidad. Para afrontar esta situación se puede emplear el modelo de fragilidad, el cual considera la distribución conjunta de los tiempos de supervivencia correlacionados, permitiendo de esa forma una estimación tanto del vínculo entre los tiempos de supervivencia como del efecto de las covariables.

El modelo de fragilidad se basa en el hecho de que la dependencia surge de un factor de riesgo común, la fragilidad, el cual comparten los individuos pertenecientes a un grupo. Dicho factor se considera como inobservable, aleatorio y con una distribución como la distribución estable positiva o la Gamma.

El modelo de fragilidad más empleado es el modelo de fragilidad compartida, el cual es una generalización del modelo de Cox. En un grupo  $i$ ,  $T_i$  es el tiempo de supervivencia para cada individuo, y su función de riesgo será (Hougaard, 2000):

$$h_i(t | h_0, \beta, \omega_i) = \omega_i h_0(t) e^{\{x_i^\top \beta\}}$$

Donde  $\omega_i$  es la fragilidad asociada al grupo  $i$ .



### 5.3. Modelos paramétricos

En los modelos paramétricos se realizan suposiciones sobre los patrones de los tiempos de supervivencia. La distribución de los tiempos de supervivencia puede representarse mediante el empleo de modelos de supervivencia paramétricos continuos, asumiéndose que, los riesgos, como función del tiempo que son, tienen una forma determinada por uno o más parámetros estimados por los datos observados. Los modelos paramétricos pueden resultar de suma utilidad y precisos en aquellas situaciones en las que los datos puedan seguir una distribución en particular.

En comparación con los modelos no-paramétricos, realizar una suposición paramétrica permite modelar una función de supervivencia con más detalle. Sin embargo, se corre el riesgo de realizar suposiciones no sustentadas por los datos.

Algunos de los modelos de supervivencia paramétricos más utilizados son:

- Distribución de Weibull
- Distribución Exponencial

Para realizar una regresión de un modelo de riesgos proporcionales, se puede emplear tanto el modelo de Weibull, como el modelo Exponencial, al tratarse este de un caso especial del modelo de Weibull. Si asumimos que la variable independiente o explicativa actúa multiplicativamente en el riesgo, empezando por la función de supervivencia del modelo de Weibull y reemplazando el riesgo base con el riesgo que incluye el efecto de la variable independiente, terminamos con una función con la forma de una distribución de Weibull, pero con diferentes parámetros. Por lo tanto, el modelo de Weibull pertenece a la familia de los modelos de riesgos proporcionales.

Puesto que ambos modelos se encuentran relacionados, puede resultar complejo decidir cual de los dos resulta el más indicado para ser empleado en cada situación, pero, mediante la estimación de  $S(t)$  empleando el estimador producto límite de Kaplan-Meier y la representación gráfica, podemos comprobar qué modelo se ajusta mejor a los datos (Rivas López y López Fidalgo, 2000):

- Modelo de Weibull: al representar  $\ln[-\ln S(t)]$  frente a  $\ln t$ , se obtiene una línea recta.
- Modelo Exponencial: al representar  $-\ln S(t)$  frente a  $t$ , se obtiene una línea recta.

#### 5.3.1. Modelo de Weibull

Corresponde con una distribución monótona del riesgo (tanto un incremento como una disminución continua del riesgo).

Para el modelo de Weibull con parámetros  $p > 0$  y  $\lambda > 0$ , la función de riesgo, la función de supervivencia y la función de densidad son:

$$h(t) = \lambda p t^{p-1}$$

$$S(t) = e^{-\lambda t^p}, \quad t \geq 0$$

$$f(t) = \lambda p t^{p-1} e^{-\lambda t^p}$$

Aquí  $\lambda > 0$  es un parámetro de proporción y  $p > 0$  un parámetro de forma, permitiendo controlar el comportamiento de la función de riesgo. Concretamente se puede observar lo siguiente (Kachman, 1999):

- $h(t)$  es monótona decreciente cuando  $p < 1$
- $h(t)$  es constante cuando  $p = 1$
- $h(t)$  es monótona creciente cuando  $p > 1$

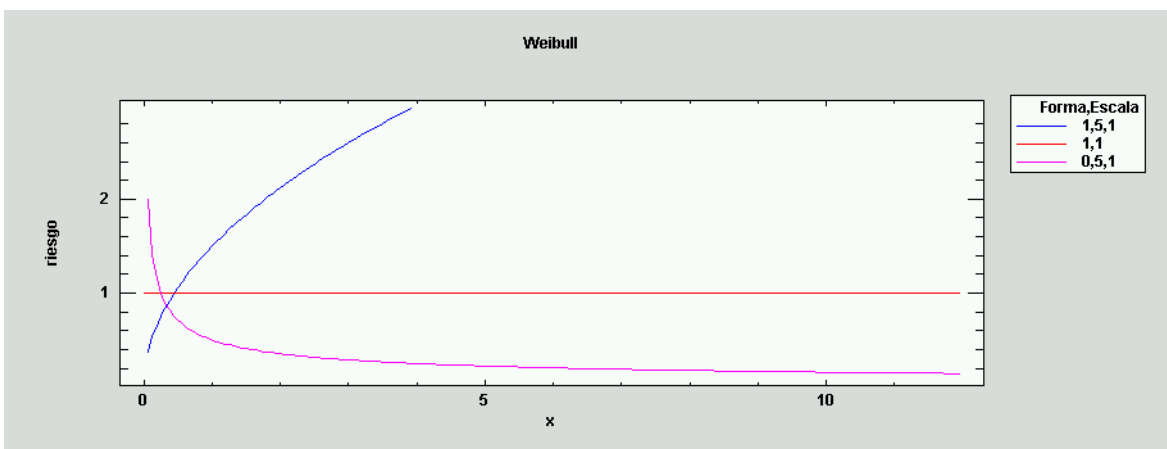


Figura 8: Representación gráfica de la función de riesgo con el modelo de Weibull, según los valores de  $p$ , obtenida mediante el programa Statgraphics.

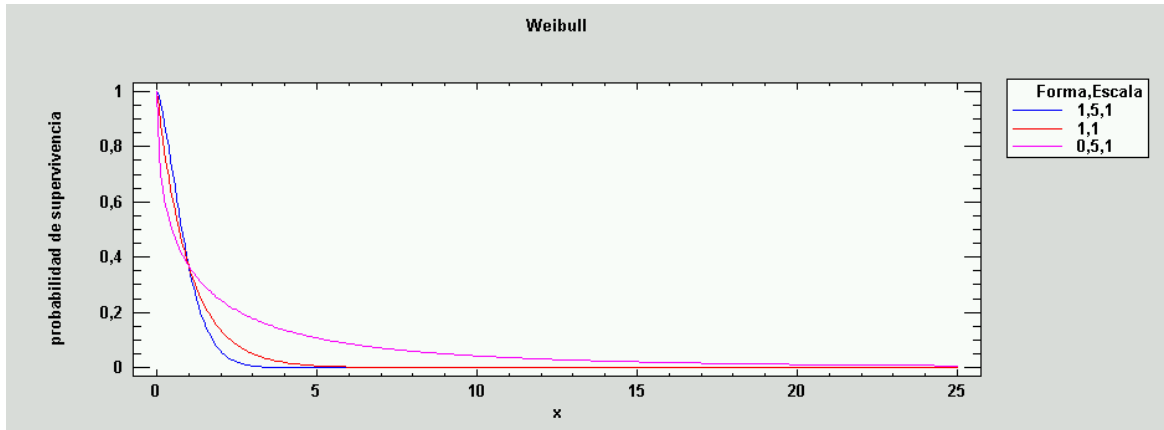


Figura 9: Representación gráfica de la función de supervivencia con el modelo de Weibull, según los valores de  $p$ , obtenida mediante el programa Statgraphics.

El tiempo de vida esperado y su varianza son:

$$E[T] = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{p}\right),$$

$$Var[T] = \frac{1}{\lambda^2} \Gamma\left(1 + \frac{2}{p}\right) - \frac{1}{\lambda^2} \Gamma\left(1 + \frac{1}{p}\right)^2$$

Otra de las propiedades del modelo de Weibull, es la linealidad de  $\ln[S(t)]$ , de modo tal que:

$$S(t) = e^{-\lambda t^p} \Rightarrow \ln S(t) = -\lambda t^p \Rightarrow \ln[-\ln S(t)] = \ln \lambda t^p = \ln(\lambda) + p \ln(t)$$

Lo que nos permite obtener  $\lambda$  y  $p$  a partir del punto de intersección y la pendiente respectivamente. (Kleinbaum y Klein, 2010)

#### Ejemplo demostrativo:

Continuando con los datos obtenidos por Freireich *et al.* (1963), empleamos el programa estadístico Statgraphics para representar gráficamente la función de supervivencia y la función de riesgo mediante una distribución de Weibull. En primer lugar, aprovecharemos la propiedad de la linealidad de  $\ln[-\ln S(t)]$  para comprobar si nuestros datos se ajustarían a un modelo de Weibull. Lo cual, en este caso se cumple (Figura 10).

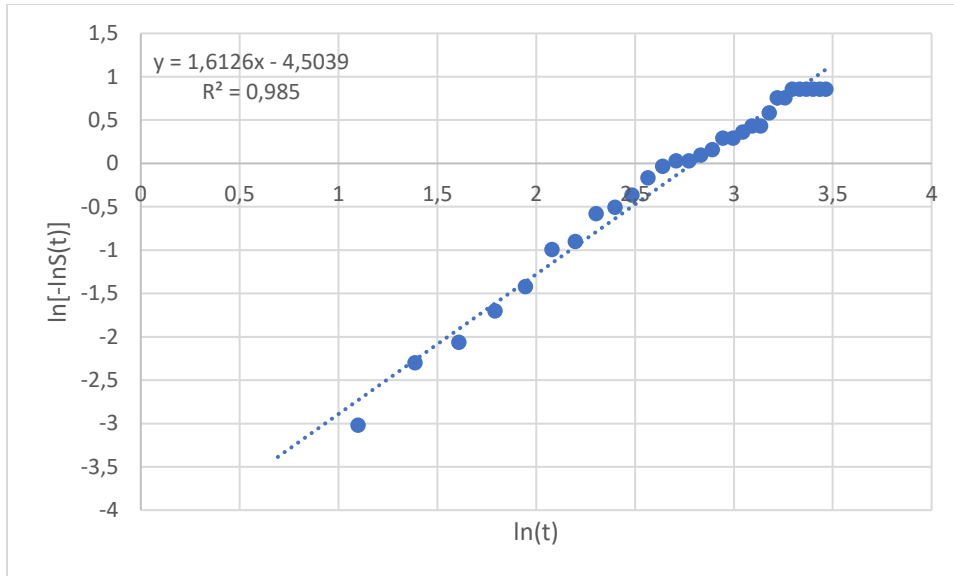


Figura 10: Representación gráfica de  $\ln[-\ln S(t)]$  con respecto al  $\ln(t)$

Como sabemos, necesitamos disponer tanto del parámetro proporción como del parámetro forma para lograr obtener el modelo de Weibull. Para ello, emplearemos el programa estadístico Statgraphics, en el cual mediante: **Describir** → **Datos de vida** → **Análisis de Weibull...**, donde estableceremos como **Datos** a TIEMPO y como **(Censurados)** a ESTADO, el método de estimación será mediante Máxima Verosimilitud y nos interesa obtener la tabla Resumen de análisis, lo que nos permitirá saber los valores de los parámetros forma y proporción (llamado Escala en Statgraphics). Los cuales en nuestro caso son  $\lambda= 17,904$  y  $p=1,14082$ , de modo tal que podemos prever que la función de riesgo será creciente.

Teniendo ya todos los datos necesarios, podemos proceder a trabajar: **Graficar** → **Distribuciones de probabilidad** → **Weibull**. Establecemos **Forma**=1,14082 y **Escala**=17,904, y seleccionamos los GRÁFICOS y TABLAS que deseamos obtener, en este caso **Función de Supervivencia** y **Función de Riesgo**. Finalmente podemos comprobar que tal y como habíamos supuesto, la función de riesgo es creciente.

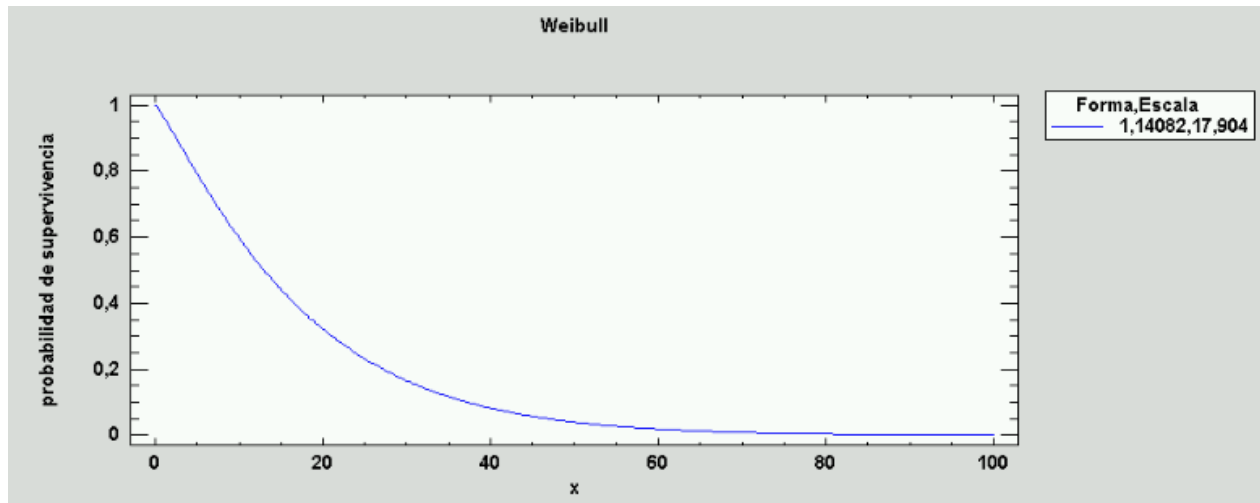


Figura 11: Representación gráfica de la función de supervivencia  $S(t)$  según la distribución de Weibull.

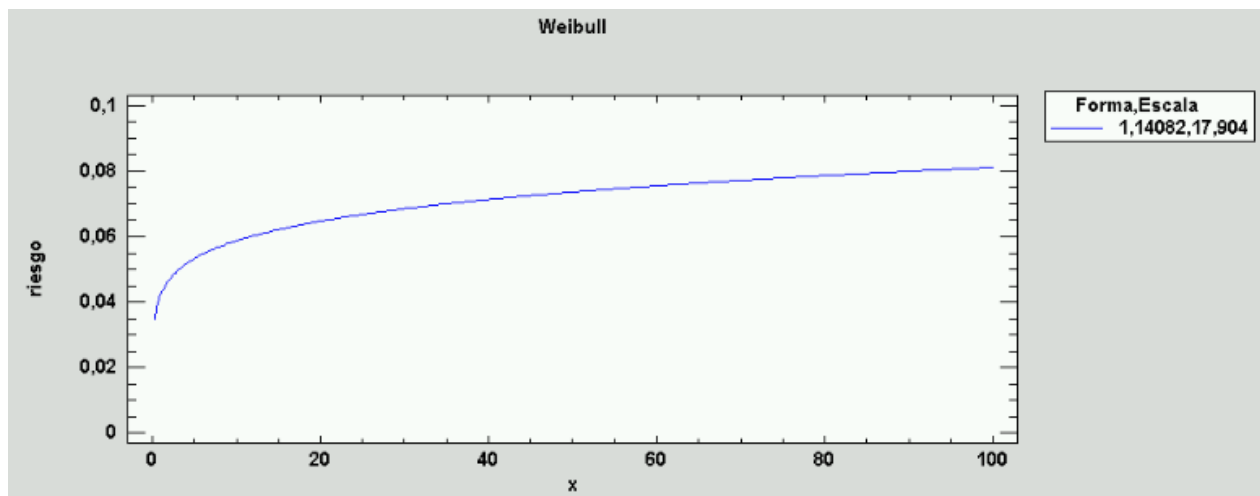


Figura 12: Representación gráfica de la función de riesgo  $h(t)$  según la distribución de Weibull.

### 5.3.2. Modelo Exponencial

Es un caso especial de la distribución Weibull que aparece cuando  $p=1$ , causando la independencia de la función de riesgo con respecto al tiempo. Es una distribución más simple puesto que solo se requiere estimar un parámetro. Cuando se emplea una distribución exponencial con unos datos dados, debe asumirse que la función de riesgo asociada es constante en el tiempo, y que, por lo tanto, la ocurrencia del evento es completamente aleatoria. A esta propiedad del modelo Exponencial, riesgo constante, se le suele llamar pérdida de memoria.

Para el modelo Exponencial, con parámetro  $\lambda$ , la función de riesgo, la función de supervivencia y la función de densidad son:

$$h(t) = \lambda$$

$$S(t) = e^{-\lambda t} \quad t \geq 0$$

$$f(t) = \lambda e^{-\lambda t}$$

El tiempo de vida esperado y su varianza son:

$$E[T] = \frac{1}{\lambda}$$

$$Var[T] = \frac{1}{\lambda^2}$$

### Ejemplo demostrativo:

Proseguimos con el programa Statgraphics y los datos de Freireich *et al.* (1963) para obtener la distribución Exponencial. Para ello, en primer lugar, debemos comprobar que dichos datos se pueden ajustar adecuadamente a un modelo Exponencial, mediante su propiedad de linealidad de  $-\ln S(t)$ . Lo que nos permite comprobar (Figura 13) que nuestros datos, a pesar de que se ajustaban al modelo de Weibull, también se pueden ajustar a un modelo Exponencial, lo cual puede ser debido al hecho de que,  $p = 1,14082 \approx 1$ , de modo tal que, aunque en condiciones normales trabajaríamos con el modelo de Weibull para obtener el resultado más fiable posible, no existe problema en emplear el modelo Exponencial a modo de ejemplo, ya que los resultados apenas se verán alterados.

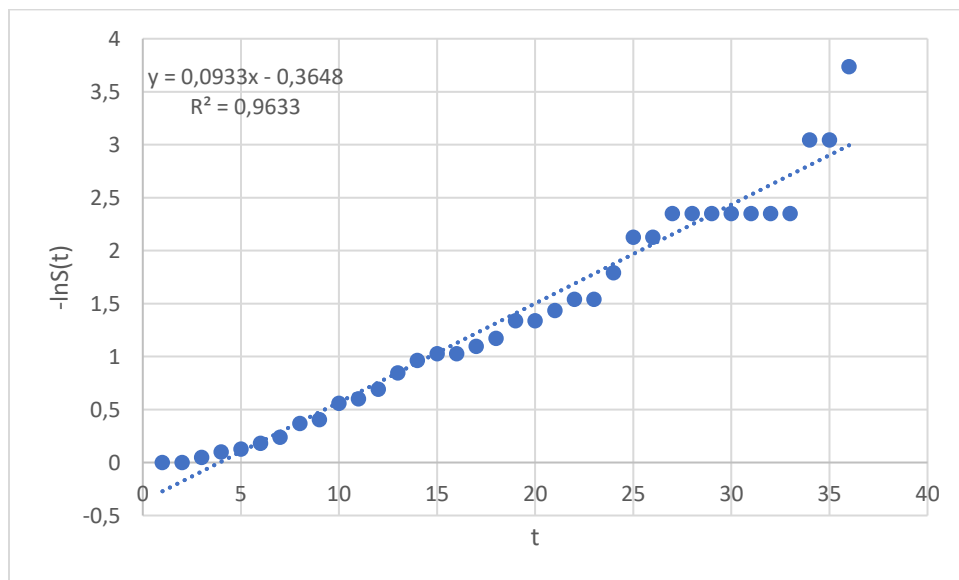


Figura 13: Representación gráfica de  $-\ln S(t)$  con respecto a  $t$ .

A continuación, con ayuda del programa Statgraphics, **Describir** → **Ajuste de distribuciones** → **Ajuste de Datos Censurados...**, donde estableceremos como **Datos** a TIEMPO y como **(Censurados)** a ESTADO, seleccionamos de entre las opciones posibles Distribución Exponencial y Resumen de Análisis, de donde obtenemos el valor de la media= 18,033. Con la media obtenida, seguimos los pasos: **Graficar** → **Distribuciones de probabilidad** → **Exponencial**. Establecemos Media= 18,033 y como Gráficos seleccionamos **Función de Supervivencia** y **Función de Riesgo**.

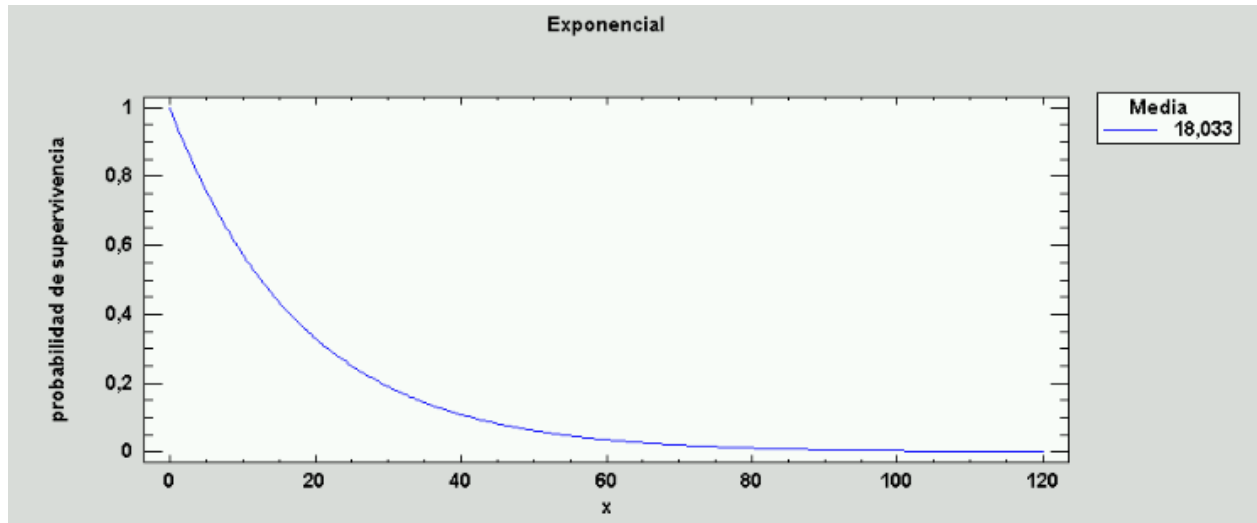


Figura 14: Representación gráfica de la función de supervivencia  $S(t)$  del modelo Exponencial.

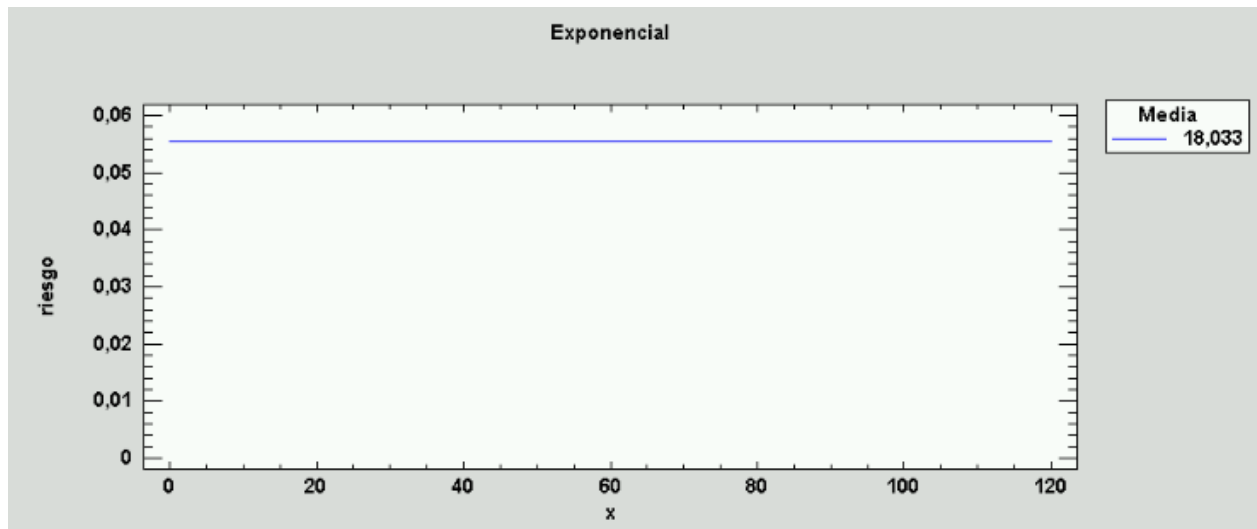


Figura 15: Representación gráfica de la función de riesgo  $h(t)$  del modelo Exponencial.

## 6. Aplicaciones

El análisis de supervivencia es una de las técnicas estadísticas más importantes en el área de la salud y en las ciencias biológicas (Kleinbaum and Klein, 2010; Collett, 2015). A pesar de que originalmente se diseñó para la investigación médica, con el tiempo sus aplicaciones fueron extendiéndose a otros dominios. Una de las claves que podrán observarse en las aplicaciones del análisis de supervivencia, es la variedad que existe en la potencial complejidad de la técnica, en algunos estudios sólo se requiere aplicar los conceptos más básicos del análisis de supervivencia (Vjollca *et al.*, 2019), mientras que en otros, determinar los factores a tener en cuenta (Alcon *et al.*, 2019), o la observación de los resultados (Lesem *et al.*, 2011) ya supone un desafío en sí mismo.

### 6.1. Medicina y Salud Humana

En el ámbito de la salud es donde más se emplean las técnicas del análisis de supervivencia. Empezando su apogeo en torno a los años 70 y continuando con su crecimiento desde entonces (Griffin y Gardner, 1989; Singer y Willet, 1991).

#### **COVID-19**

Con la propagación en 2019 de una enfermedad respiratoria potencialmente mortal hasta alcanzar el estado de pandemia, toda la información que se pudiera recabar con respecto a dicha enfermedad pasó a ser considerada como prioritaria para hacerle frente de la manera más eficiente posible, y tratando de reducir al mínimo el número de víctimas mortales que esta pudiera ocasionar.

El análisis de supervivencia ha demostrado su valía en este momento observándose como el número de estudios relacionados con la COVID-19 y el análisis de supervivencia se incrementa de manera continua. De entre esos estudios, puede mencionarse el realizado por Dasgupta y Choudhury (2021), en el cual analizaron la relación del género y la edad con el tiempo de permanencia necesario en el hospital, tratando de esa manera de ayudar a que los servicios médicos puedan organizar sus equipos e instalaciones de forma más eficiente. Para llevarlo a cabo, durante 40 días analizaron mediante el método de Kaplan-Meier los datos relacionados con el género de los pacientes (Figura 16), lo que, junto al modelo de regresión de Cox permitió determinar que el género no afectaba a la supervivencia ni al tiempo de estancia de los pacientes de manera significativa.



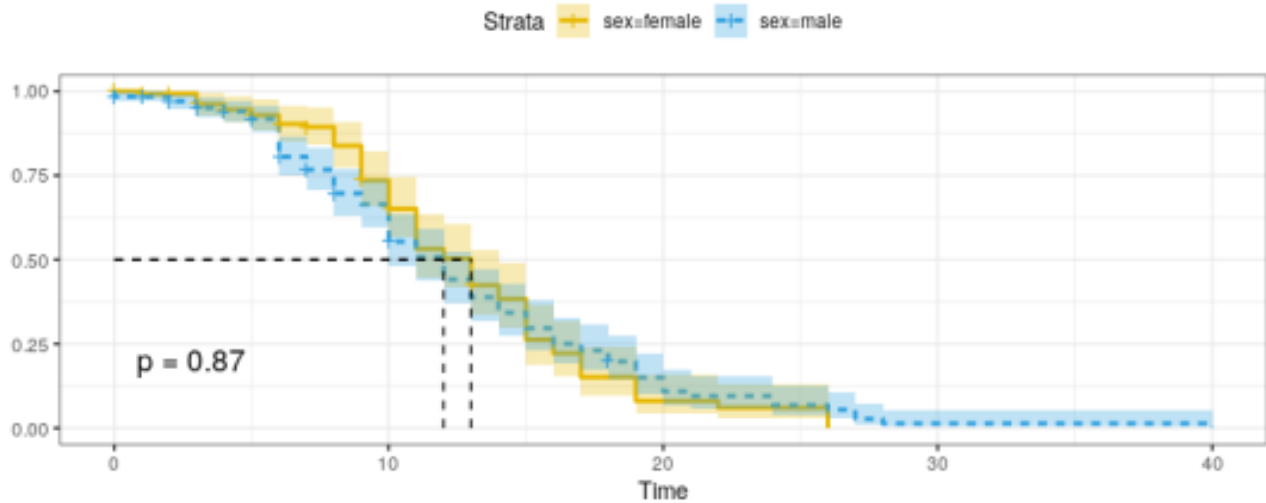


Figura 16: Representación gráfica de las curvas de Kaplan-Meier donde se observa la comparación de la supervivencia entre los pacientes agrupados por su género.

Posteriormente, se llevó a cabo el mismo análisis, solo que la agrupación de los pacientes se realizó por edad, concretamente mediante el rango de percentiles, de modo tal que se dividieron en tres categorías:

- Categoría 1: edad menor al percentil 25
- Categoría 2: edad superior al percentil 25 pero inferior al percentil 75
- Categoría 3: edad superior al percentil 75

Esta vez, los resultados obtenidos mediante el método de Kaplan-Meier mostraron que el tiempo de recuperación y la probabilidad de supervivencia se veían afectados de manera significativa por la edad de los pacientes (Figura 17) comprobándose que aquellos individuos de mayor edad pertenecientes a la categoría 3, tenían una menor probabilidad de supervivencia con respecto a los de menor edad pertenecientes a la categoría 1 y a la categoría 2.

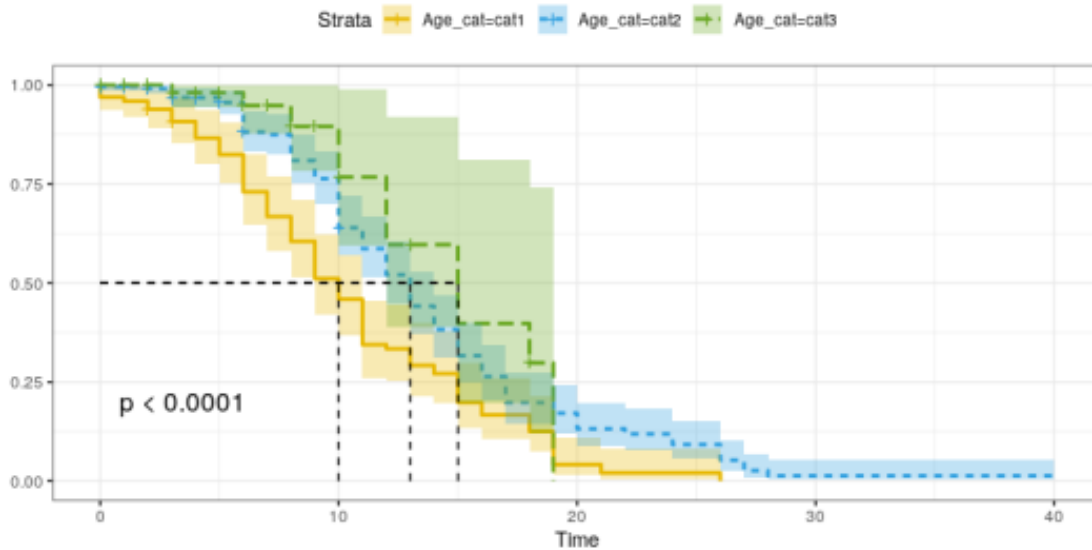


Figura 17: Representación de las curvas de supervivencia en las que se compara la probabilidad de supervivencia de los pacientes según su edad.

Otro ejemplo sería el estudio realizado por Emami *et al.* (2021), donde estudia la influencia de poseer una comorbilidad como la hipertensión en los pacientes infectados por COVID-19. El estudio se realizó entre el 20 de febrero y el 1 de marzo, con un total de 1239 pacientes a los que se dividió según tuvieran o no hipertensión (Figura 18).

Variables	Total (1239)	Hypertension (159)	Non-hypertension (1080)	P-value
Age, years	51.48 ± 19.54	66.18 ± 12.70	48.95 ± 19.40	P < .001
Sex				
Male	692 (55.9%)	72(45.30%)	620 (57.4%)	0.003
Female	547 (44.1)	87(54.7%)	460 (42.6%)	
Symptoms				
Fever	459 (37%)	52 (32.7%)	407 (37.7%)	0.13
Cough	614 (49.6%)	90 (56.6%)	524 (48.5%)	0.03
Muscle ache	257 (20.7%)	31 (19.5%)	226 (20.9%)	0.38
Diarrhea	6 (0.5%)	-	6(0.5%)	0.43
Headache	11 (0.9%)	4 (2.5%)	7 (0.6%)	0.04
Chest Pain	8 (0.6%)	2 (1.3%)	6 (0.6%)	0.27
Dizziness	7 (0.6%)	1 (14.3%)	6 (85.7%)	0.61
Respiratory Distress	576 (46.5%)	94 (59.1%)	482 (44.6%)	P < .001
PO2 < 93%	481 (38.8%)	86 (54.1%)	395 (36.6%)	P < .001
Abnormal Chest CT	421 (34%)	79 (18.8%)	342 (81.2%)	P < .001
Comorbidities				
Diabetes mellitus	176 (14.2%)	68(42.76%)	108 (10.0%)	P < .001
Cardiovascular disease	132 (10.7%)	45 (28.3%)	87 (8.1%)	P < .001
Chronic liver disease	22 (1.8%)	4 (2.5%)	18 (1.7%)	0.309
Chronic kidney disease	44 (3.6%)	15 (34.1%)	29 (65.9%)	P < .001
Cancer	22 (1.8%)	2 (1.3%)	20 (1.9%)	0.44
HIV	5 (0.4%)	0 (0%)	5 (0.5%)	0.503
Smoking	27 (2.5%)	8 (5.2%)	19 (2.0%)	0.02
Addict	23 (2.1%)	4 (2.6%)	19 (2.0%)	0.41
Asthma	25 (2.0%)	8 (5.0%)	17 (1.6%)	0.01
Immunodeficiency Disease	10 (0.8%)	1(0.6%)	9 (0.8%)	0.62
Clinical prognosis				
ICU admission	98 (7.9%)	16 (10.06%)	82 (83.7%)	0.09
Intubation	106 (8.55%)	19 (11.94%)	87	0.24
Death	98	16	82	0.09
Length of hospital stay	5.44 ± 6.89	6.60 ± 6.61	5.27 ± 6.92	0.02
Onset of symptom to hospital admission	5.62 ± 5.1	4.82 ± 2.28	5.00 ± 3.25	0.41

Figura 18: Características clínicas de los pacientes con hipertensión y sin hipertensión.

Para obtener la función de supervivencia de los dos grupos con respecto al tiempo, se emplearon el método de Kaplan-Meier y la prueba log-rank, mientras que para determinar los riesgos relativos a otras comorbilidades distintas a la hipertensión emplearon la regresión de Cox. De acuerdo con los resultados obtenidos por la curva de Kaplan-Meier y la prueba log-rank, los pacientes con hipertensión empeoraban más rápidamente que aquellos sin hipertensión (Figura 19), pero, al mismo tiempo, observaron una relación entre aquellos pacientes con hipertensión y la edad, teniendo generalmente mayor edad aquellos con hipertensión, factor que también afecta al nivel de gravedad de la enfermedad.

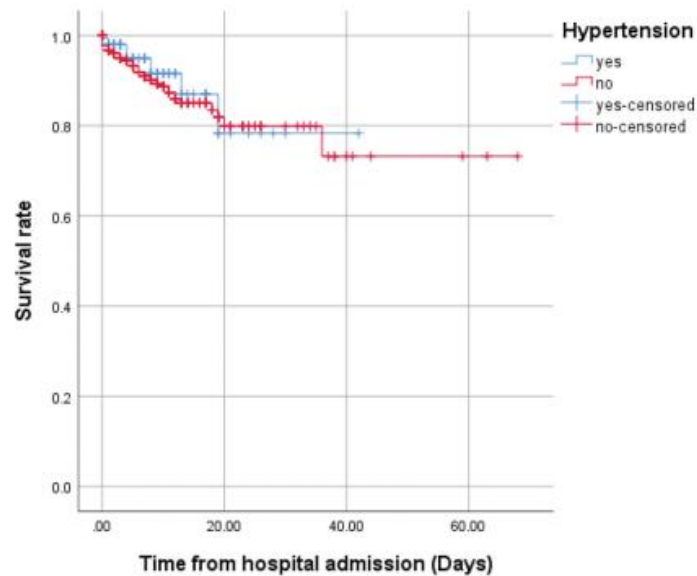


Figura 19: Curva de Kaplan-Meier en la que se compara la probabilidad de supervivencia de los pacientes con hipertensión y sin hipertensión.

Mediante el modelo de regresión de los riesgos proporcionales de Cox, se pudo determinar que aquellos pacientes con enfermedades cardíacas, hipertensión, VIH y enfermedades renales tienen una mayor probabilidad de muerte que aquellos sin comorbilidades.

Variables	HR (95%CI)	p-value
Age	1.41 (1.23–2.5)	0.02**
Diabetes mellitus	1.25 (0.36–4.36)	0.72
Cardiovascular disease	3.52 (1.23–11.15)	0.02**
Chronic liver disease	0.807 (0.35–1.15)	0.807
Chronic kidney disease	1.15 (0.35–3.74)	0.012**
Cancer	0.58 (0.08–4.24)	0.59
HIV	10.46 (3.27–33.45)	P < .001**
Smoking	0.61 (0.15–2.52)	0.5
Asthma	1.04 (0.53–2.02)	0.89
Immunodeficiency Disease	1.65 (0.98–1.42)	0.97

Figura 20: Regresión de Cox multivariante con riesgos relativos a comorbilidades y COVID-19.

### Efecto de la quimioterapia en pacientes con cáncer de mama

Liedtke *et al.* (2008) estudió el efecto de la quimioterapia en pacientes con cáncer de pecho triple negativo (TNBC) el cual se caracteriza por la ausencia de expresión de tres genes: receptor de estrógeno, receptor de progesterona y el receptor 2 del factor de crecimiento epidérmico humano.

A los 1.118 pacientes a los que se sometió al estudio se les dividió en pacientes TNBC y pacientes sin TNBC, siendo un 23% y un 77% respectivamente, para posteriormente medir el tiempo desde la cirugía hasta la muerte, o la finalización del seguimiento.

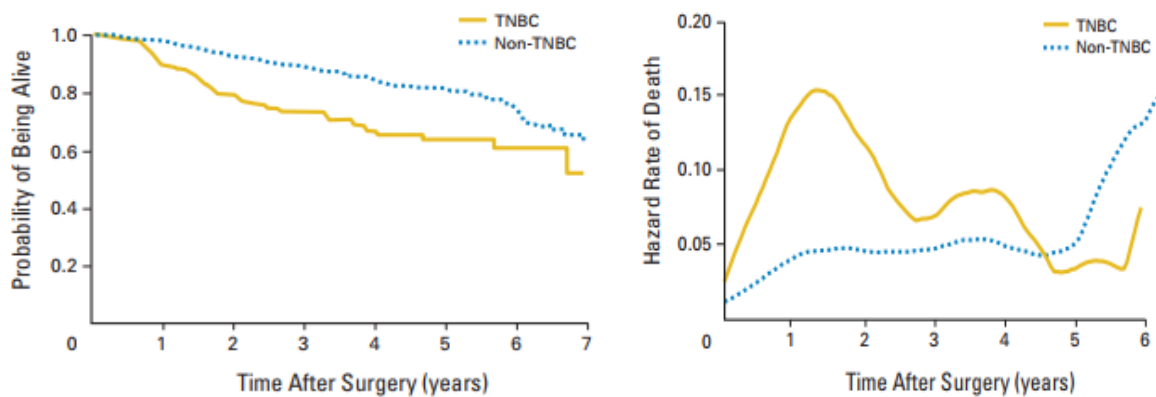


Figura 21: Representación gráfica de la función de supervivencia y la función de riesgo en las que se compara a los pacientes TNBC con los TNBC.

De esta forma, los autores encontraron que los pacientes con TNBC poseen una supervivencia menor que aquellos que no lo tienen.

### Efecto de la medicación en pacientes con esquizofrenia

Lesem *et al.* (2011) estudiaron el efecto de medicamentos en individuos con esquizofrenia. Debido a la complejidad de este trastorno neurológico, resulta difícil juzgar únicamente observando a dichos individuos cuanto tiempo dura la medicación o el comienzo de un ataque. Por lo que, medir el tiempo hasta un ataque no aportaba información al no ser directamente observable.

Con objeto de lograr llevar a cabo el estudio, los autores seleccionaron un total de 344 pacientes, tanto a hombres como mujeres de entre 18 y 65 años, diagnosticados con esquizofrenia, y se los dividió en tres grupos según lo que se les administró: placebo, 5 mg de loxapina y 10 mg de loxapina.

En un momento dado, los pacientes usaron un inhalador que les administró la medicación. Durante las dos horas siguientes, no se les permitió utilizar nuevamente el inhalador. A partir de esas dos horas se les permitió usarlo nuevamente, empleando ese tiempo entre los dos usos como unidad de medida del tiempo a modo de tiempo hasta el evento, que pudo emplearse para realizar un análisis de supervivencia para evaluar las diferencias entre los medicamentos.

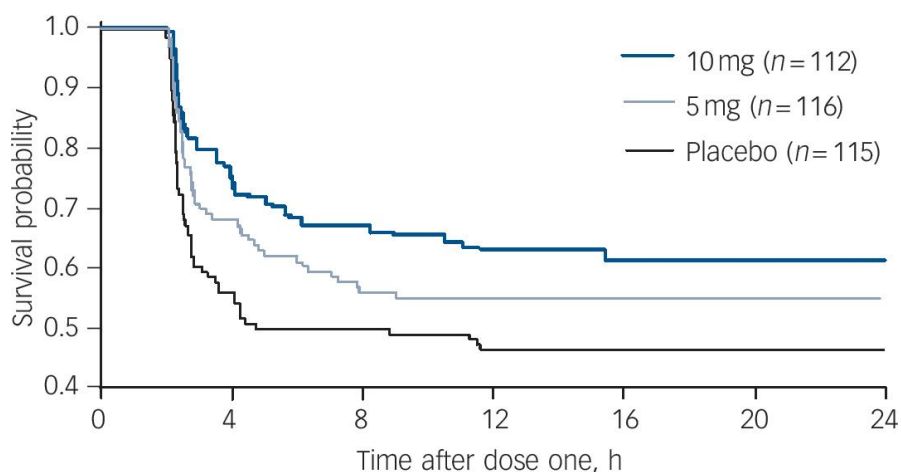


Figura 22: Análisis de supervivencia mediante una curva Kaplan-Meier que representa el tiempo hasta la segunda dosis.

De esta forma descubrieron que la loxapina inhalada proporcionaba una reducción significativa en la agitación de los pacientes en comparación con el placebo. Esta investigación sirve como ejemplo para mostrar la dificultad para establecer una escala de tiempo en las investigaciones de complejidad subyacente.

#### Búsqueda de tratamiento para la malaria

Moreira *et al.* (2021) investigaron como afectaba la administración de dexametasona en la supervivencia, el desarrollo de la parasitosis y el estado redox de ratones (*Mus musculus*) infectados con *Plasmodium berghei*, el cual posee un cuadro clínico y un genoma similar a *Plasmodium falciparum*, causante de la malaria en el ser humano. Para llevar a cabo el estudio, dividieron a los ratones en distintos grupos:

- Control positivo: inoculados con *P. berghei* pero sin tratamiento.
- Grupo con tratamiento de dexametasona e inoculados con *P. berghei*.
- Grupo con tratamiento de L-arginina e inoculados con *P. berghei*.
- Control negativo: sin tratamiento y sin infección.

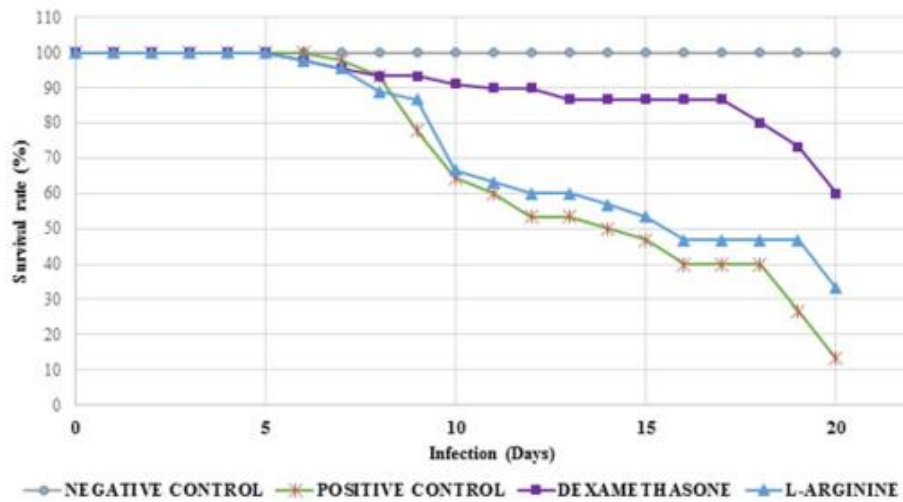


Figura 23: Curvas de supervivencia donde se comparan los distintos tratamientos empleados en los ratones inoculados.

Los resultados mostraron una inhibición de la síntesis de óxido nítrico en el cerebro de aquellos ratones a los que se trató con dexametasona, incrementando de esa manera su tasa de supervivencia frente a *P. berghei* (ver Figura 23).

### Descubrimiento de subtipos de cáncer

Bair y Tibshirani (2004) combinaron datos sobre la expresión de genes y datos clínicos, puesto que consideraban un problema identificar los subtipos de cáncer sin poseer información de los distintos subtipos de cáncer en la base de datos.

En primer lugar, emplearon la regresión de Cox para medir la correlación entre la expresión de genes y la supervivencia de los pacientes, para, de esa forma, seleccionar únicamente aquellos genes que obtuvieran una puntuación superior a 2,39. Una vez que los seleccionaron, procedieron a agruparlos mediante técnicas de clustering tradicionales, como K-means. Posteriormente, probaron la asignación de los grupos utilizando únicamente los datos clínicos. Empleando los datos clínicos, establecieron la asignación de grupos como la variable dependiente y aplicaron algoritmos de clasificación.

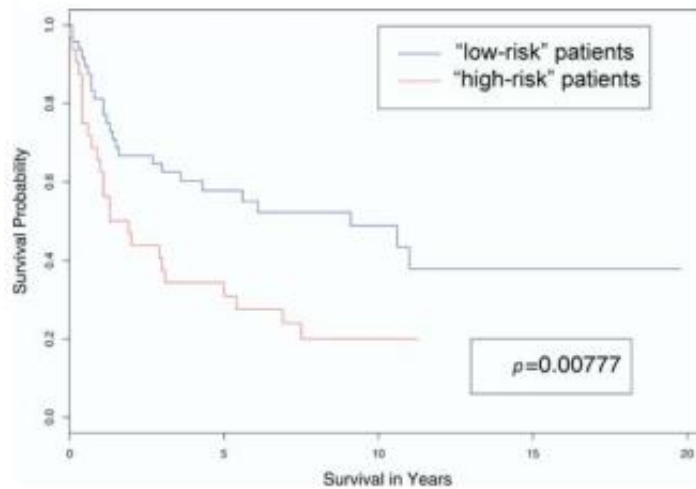


Figura 24: Comparación de las curvas de supervivencia obtenidas para los grupos de pacientes con p-valor=0,0777 en la prueba log-rank.

Finalmente, pudieron comprobar que los algoritmos de clasificación funcionaron correctamente, lo que implicaba que la asignación de los grupos fue identificada de manera exitosa.

## 6.2. Vegetal

### Supervivencia de variedades silvestres de trigo

Las variedades silvestres de aquellas plantas domesticadas por el ser humano resultan de especial importancia como fuente de variedad genética y como base para la mejora genética de nuevos cultivos. Por ello, Vjollca *et al.* (2019) estudió como afectaba el estrés hídrico a ocho variedades de *Aegilops* distintas, observando la germinación de las semillas a las 24 y 72 horas, y el porcentaje de plántulas que sobreviven una vez expuestas a estrés hídrico.

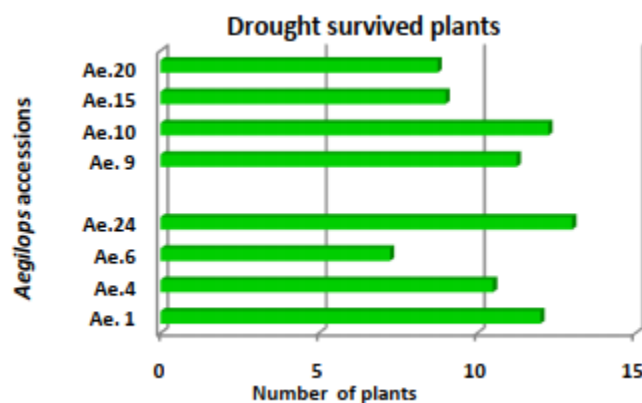


Figura 25: Número de plantas de cada variedad que sobrevivieron al finalizar el periodo de estudio

Posteriormente se estudió la correlación entre el porcentaje de supervivencia de las plántulas con el porcentaje de germinación, determinando así cuales manifestaban una mayor tolerancia al estrés, convirtiéndose en potenciales candidatas para la mejora vegetal.

### Adquisición de tecnología de irrigación

La investigación llevada a cabo por Alcon *et al.* (2019) resalta los niveles de complejidad que puede llegar a alcanzar un análisis de supervivencia que involucra al ser humano en un tema aparentemente tan simple, como puede ser cambiar la irrigación convencional de los cultivos, por el sistema de riego por goteo. Para que el estudio pudiera llevarse a cabo, se tuvo en cuenta una gran variedad de factores para el análisis:

- Características del granjero y de la granja: El nivel de educación, la edad, pertenencia a cooperativas involucradas en el tema, tamaño de la granja, tipos de cultivo.
- Factores económicos: tamaño del negocio, estado económico, porcentaje de los ingresos que provienen de la granja, precio del agua.
- Factores institucionales: fuente de información sobre la tecnología, posibilidad de probar la tecnología antes de probarla.
- Tecnología: expectativas generadas por la nueva tecnología, potencial de mejora por su adquisición, nivel de complejidad de la tecnología.

El análisis de supervivencia en este caso se realizó considerando la transición al nuevo sistema de riego como el evento (fallo) y permanecer con el sistema tradicional como supervivencia, por lo tanto, la probabilidad de que un granjero cambiase al nuevo sistema de riego era el riesgo.

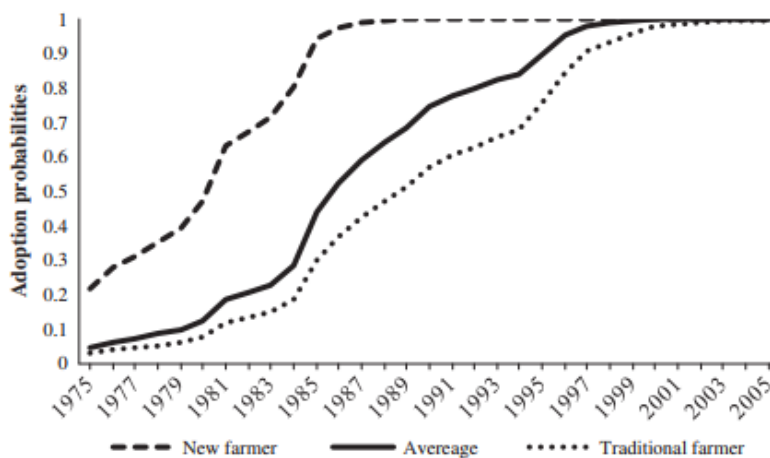


Figura 26: Representación gráfica de la probabilidad de adquirir el nuevo sistema de riego, en la que se comparan las curvas de tres grupos de granjeros, a los que se les asignó una lista de parámetros.



Se agrupó como censurados por la derecha a aquellos que decidieron realizar el cambio; a los que ya habían adquirido el sistema de manera previa al estudio como censurados por la izquierda y por último a los que aún no lo habían adquirido. Como modelo se empleó el modelo de Weibull, para permitir de esa manera la dependencia de la función con respecto al tiempo al mismo tiempo que se empleaba una función paramétrica. Los resultados mostraron la importancia de la disponibilidad de la información tecnológica, especialmente en aquellas áreas donde debe cuidarse el suministro de agua; otros factores como la pertenencia a una cooperativa, la posibilidad de probar la tecnología o la disponibilidad económica favorecen el cambio.

### Supervivencia de injertos

García (2012) estudió las causas de mortalidad de injertos en *Caryodendron orinocense*, puesto que sus propiedades como planta oleaginosa le otorga gran valor nutricional e industrial. Para llevarlo a cabo, evaluó el tiempo de supervivencia de los injertos, considerando como fallo la muerte del injerto. Los factores que tuvo en cuenta fueron el sexo de la planta donadora, el tiempo que se tardó en quitar el plástico que cubría los injertos, y la dosis de nitrógeno que se aplicaba.

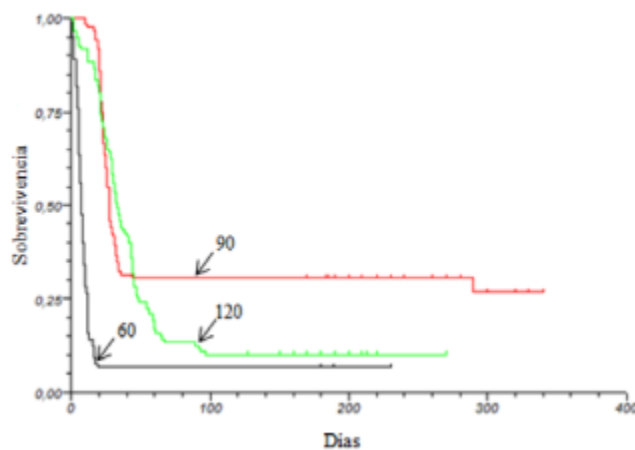


Figura 27: Curva de supervivencia en la que se comparan los efectos de quitar el plástico en distintos momentos.

Para interpretar los resultados, se aplicó el modelo de Kaplan-Meier y el modelo de Cox, llegando a la conclusión de que el sexo no influía en la supervivencia, en cambio, las dosis de nitrógeno y el tiempo de permanencia del plástico sí, determinándose de esa manera la dosis y el tiempo óptimos.

	N=9g.p <sup>-1</sup>			N=12 g.p <sup>-1</sup>			N= 18g.p <sup>-1</sup>		
	D60	D90	D120	D60	D90	D120	D60	D90	D120
Muertes observadas	39	30	38	36	19	33	37	35	37
Muertes esperadas	10,14	46,83	50,02	13,20	38,44	36,36	16,04	47,06	45,90
Riesgo relativo (RR)	3,85	0,64	0,75	2,73	0,49	0,91	2,31	0,74	0,81
Log-Rank	$\chi^2$ para RR = 108,62 P<0,0001			$\chi^2$ para RR = 54,49 P<0,0001			$\chi^2$ para RR = 36,07 P<0,0001		

g.p<sup>-1</sup> = g planta<sup>-1</sup>/aplicación y D60, D90 y D120 = desprendimiento del plástico que cubre el injerto a los 60, 90 y 120 días de la injertación, respectivamente.

Figura 28: Resultados obtenidos mediante los modelos de Kaplan-Meier y Log-rank en las que se consideran la dosis de nitrógeno y el tiempo hasta que se retira el plástico.

### 6.3. Animal

El análisis de supervivencia tiene también diversas aplicaciones en el área animal, enfocándose tanto en la agricultura, como en la conservación de las especies.

#### Prevención de septicemia en peces criados para acuicultura

Como en el estudio realizado por Yazdanpanah-Goharrizi *et al.* (2020) donde el análisis de supervivencia no se aplicó sobre el animal de interés directamente, que en este caso serían los peces, en su lugar, se realizó sobre *Aeromonas hydrophila*, principal causante de la septicemia en peces, para buscar una manera de eliminar o al menos controlar la presencia de dicha bacteria mediante el uso de bacteriófagos, logrando de esa manera demostrar la utilidad del bacteriófago AH $\phi$ 3 como herramienta de control biológico.

#### Efecto de nuevos insecticidas en las abejas

Las abejas se encuentran entre los insectos polinizadores, y, a pesar de su utilidad para la polinización de plantas y cultivos, su número se está viendo reducido. Por ello, se investigó (Boff *et al.*, 2021) el efecto que tiene en la supervivencia de *Osmia bicornis* la exposición a largo plazo a distintas dosis de Sulfoxaflor. El experimento se supervisó y censuró dos veces al día durante cinco días, realizando el análisis de supervivencia empleando el estimador de Kaplan-Meier y realizando comparaciones por parejas mediante la prueba log-rank. Después de cinco días, de 85 abejas que había al inicio del estudio, 41 sobrevivieron, no observándose ninguna diferencia en la supervivencia de las abejas del grupo control y aquellas suministradas con 5 y 10  $\mu\text{g dm}^{-3}$  de Sulfoxaflor, en cambio, el grupo al que se administró 50  $\mu\text{g dm}^{-3}$  vio su probabilidad de supervivencia reducida cerca de  $\frac{3}{4}$  de su valor original (Figura 29).

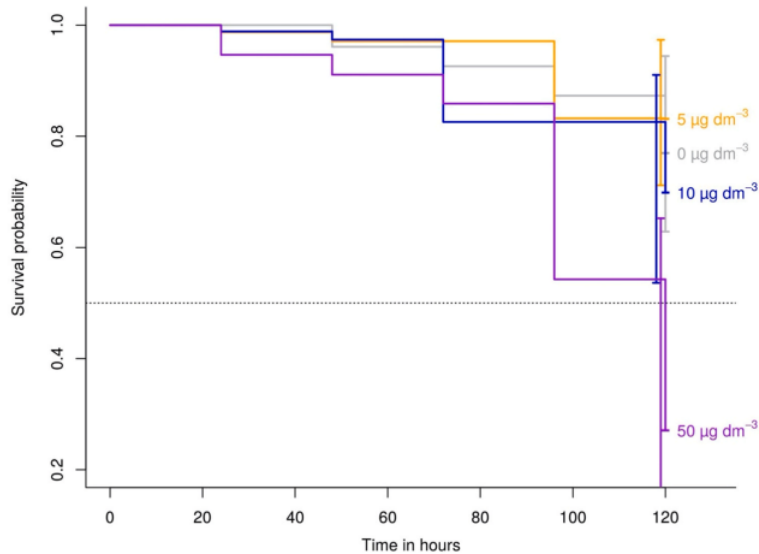


Figura 29: Efecto de las distintas dosis de Sulfoxaflor en la probabilidad de supervivencia de *Osmia bicornis* con respecto al paso de las horas desde su suministración.

Al final se determinó que la dosis de insecticida que habitualmente se emplea en los campos, no solo tenía un impacto negativo en la supervivencia de *Osmia bicornis*, sino que además afectaba a su conducta y su capacidad de polinizar.

### Supervivencia y reproducción de ratas de laboratorio

*Rattus norvegicus*, coloquialmente conocida como rata de laboratorio, resulta de suma importancia en el desarrollo de investigaciones de todo el mundo, llegando a emplearse millones de individuos al año. Debido a su papel y relevancia, Abu *et al.* (2020) llevaron a cabo un estudio para determinar cómo afectaba la temperatura al índice de supervivencia y la reproducción de estos roedores, para determinar así los momentos del año en los que es más idónea su reproducción y en cuales requieren de especial cuidado para evitar el estrés por calor. Para ello, tuvieron en cuenta los valores tanto del porcentaje de nacimientos como del porcentaje de muertes a lo largo de cinco meses, llegando a la conclusión de que existe una correlación directa entre el aumento de la temperatura y el porcentaje de ratas muertas (Figura 30), al mismo tiempo que existe una correlación inversa entre el aumento de la temperatura y el porcentaje de nacimientos (Figura 31).

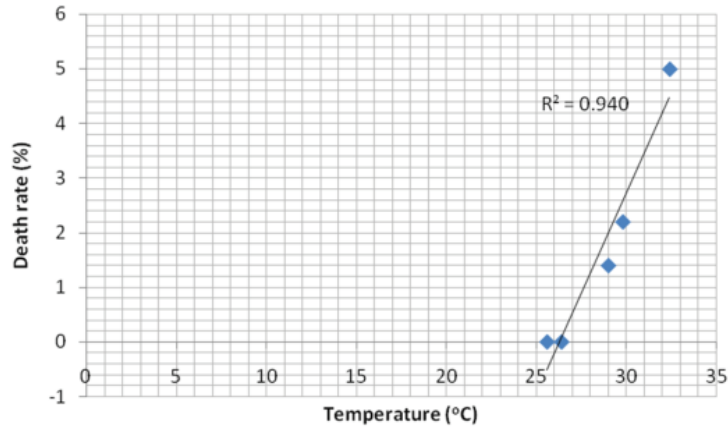


Figura 30: Correlación entre la temperatura y el índice de muerte de ratas de laboratorio.

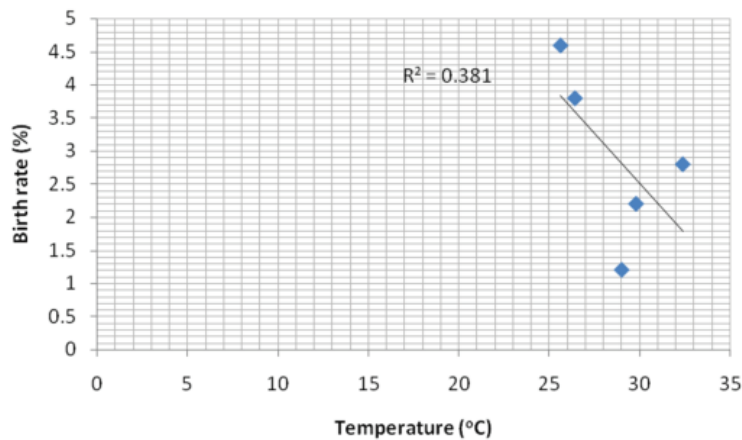


Figura 31: Correlación entre la temperatura y el índice de nacimiento de ratas de laboratorio.

### Supervivencia de aves protegidas

Se estudió (Robinson *et al.*, 2019) como algunos factores como la dieta, podían afectar a la supervivencia de *Haematopus bachmani*, un ave protegida por su pequeña población y escasa distribución. Para llevar a cabo el estudio, se tuvieron en cuenta no solo el aporte energético de su dieta, también consideraron la temperatura, la edad y la localización del nido. El monitoreo se realizó cada 3-5 días desde que salieran del huevo, hasta lograr alzar el vuelo (pasando a ser considerado como censura por abandonar la zona de estudio) o murieran (fallo). Al finalizar el estudio determinaron que la edad y el consumo de calorías eran los parámetros que más influían en la supervivencia, mostrando la relación entre edad y supervivencia una función cuadrática (Figura 32) lo cual se hipotetizó que tenía relación con el hecho de que, con el paso de los días, el aumento de tamaño facilitaba su adaptación al clima, pero, al mismo tiempo, se vuelven más activos y notables ante los depredadores.

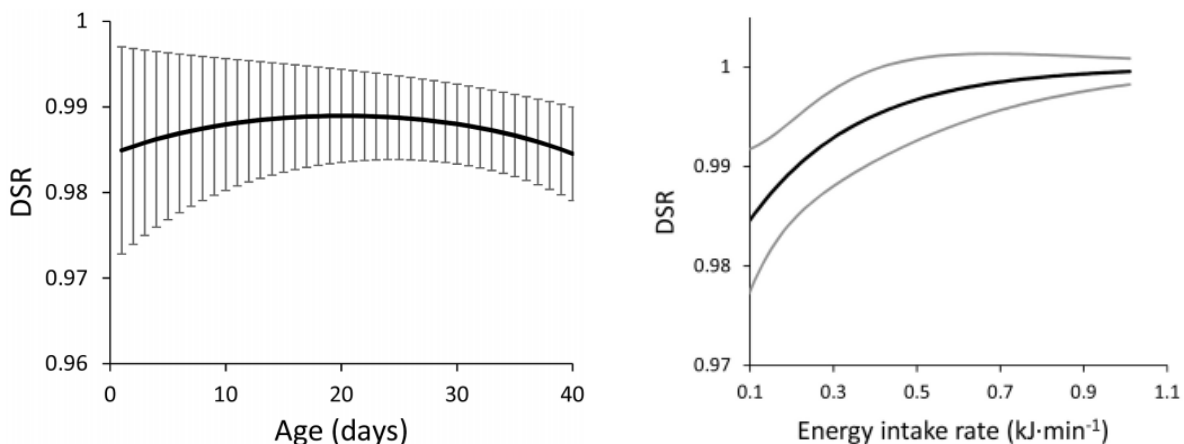


Figura 32: Índice de supervivencia diario (DSR) de las aves según su edad y su obtención de calorías.

### Mejora de la supervivencia de los terneros

A pesar de que a nivel mundial la probabilidad de supervivencia de los terneros es alta, se ha observado que, en algunos países como China, su probabilidad de sobrevivir es levemente menor, alcanzándose una diferencia de incluso un 10% con otros países como Australia. Huang (2018) resaltó la importancia de la alimentación en la salud y el crecimiento de los terneros, determinando como posible causa las técnicas de alimentación más desarrolladas, llamadas “calf kitchen” por su parecido a la cocina humana.

### Índice de supervivencia de lechones y bienestar animal

Buoio y Costa (2020) analizaron la supervivencia de los lechones durante la lactancia, puesto que aproximadamente el 20% mueren aplastados por las madres. Para ello, decidieron comparar cuales serían los resultados de una jaula de maternidad convencional, con los de una jaula especialmente pensada para el bienestar del animal. Se tuvo en cuenta cuantos nacieron vivos, cuantos nacieron muertos, cuantos fueron aplastados en distintos periodos de tiempo y cuantos sobrevivieron hasta el final.

Piglets	WFC	SEM	CFC	SEM
Born alive	11.71	0.21	11.95	0.23
Born dead	1.27	0.08	0.91	0.09
Mummies	0.35	0.05	0.25	0.06
Crushed at d 1	0.20	0.03	0.15	0.03
Crushed from d1 to d 3	0.57	0.06	0.48	0.06
Crushed from d 3 to weaning	0.40	0.04	0.32	0.04
Total crushed	1.17	0.08	0.95	0.08
Weaned	9.99	0.08	10.51	0.08

Figura 33: Comparación de los resultados obtenidos a partir de los dos tipos de jaula de maternidad

Curiosamente, los resultados mostraron un incremento en el confort de la madre, en contraposición a una menor tasa de supervivencia de los lechones, resultando en un mayor porcentaje de aplastamiento de

estos, denotando la necesidad de profundizar en el estudio de un sistema que permita tanto asegurar el bienestar de los animales, como su seguridad.

#### **6.4. Ambiental**

Aunque es uno de los campos en los que menos se ha empleado, el análisis de supervivencia puede aplicarse incluso en el campo de la biotecnología ambiental. Estudiando entre otras cosas:

- Los efectos que tiene la temperatura en la supervivencia de las hormigas asociadas a las áreas urbanas, ayudando a comprender como afectaría el calentamiento global a su supervivencia (Praxedes *et al.*, 2019).
- Como afectaría la implementación de las energías renovables marinas a su respectivo ecosistema y a las especies autóctonas, para evitar de esa manera un impacto negativo en estos, al mismo tiempo que se logra un desarrollo sostenible (Daunt *et al.*, 2020).
- La evaluación de la supervivencia de los manglares en distintas áreas como indicador ambiental del nivel de salinidad y contaminación (Mitra *et al.*, 2017).

### **7. Conclusiones**

- El análisis de supervivencia manifiesta múltiples aplicaciones y ha demostrado su utilidad en diversas áreas de la biotecnología, en especial en el área de la salud.
- El desarrollo de las nuevas tecnologías como el Machine Learning pueden llevar al análisis de supervivencia a un nuevo nivel de desarrollo al permitirle recopilar y emplear con mayor facilidad grandes volúmenes de datos.

## Bibliografía

- Aalen, O. (1978) Nonparametric inference for a family of counting processes. *Ann. Stat.* 6, 701–726.
- Abu, M. S., Yakubu, O. E., Udeh, S. M. C., Imadojemu, O. G., & Ogar, R. A. (2020). Effects of Heat Stress on Reproduction and Survival Rate of Wistar Rats. *J Biochem Analyt Stud*, 4(2).
- Akbar, A., Pasha, G. R., & Naqvi, S. F. H. (2009). Properties of Kaplan-Meier estimator: group comparison of survival curves. *European Journal of Scientific Research*, 32(3), 391-397.
- Alcon, F., de Miguel, M. D., & Burton, M. (2011). Duration analysis of adoption of drip irrigation technology in southeastern Spain. *Technological Forecasting and Social Change*, 78(6), 991-1001.
- Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol*, 2(4): e108
- Boff, S., Scheiner, R., Raizer, J., & Lupi, D. (2021). Survival rate and changes in foraging performances of solitary bees exposed to a novel insecticide. *Ecotoxicology and Environmental Safety*, 211, 111869.
- Buoio, E., & Costa, A. (2020). Space allowance and piglets survival rate in the farrowing crate. *Large Animal Review*, 26(5), 239-247.
- Collett, D. (2015). Modelling survival data in medical research. CRC press.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- Daunt, F., Fang, Z., Howells, R., Harris, M., Wanless, S., Searle, K., & Elston, D. (2020). Improving estimates of seabird body-mass survival rates. *Scottish Marine and Freshwater Science Vol 11 Nº 13*
- Dasgupta, S., & Choudhury, S. R. (2021). Surviving COVID 19-a survival analysis approach. *Journal of Huazhong University of Science and Technology ISSN*, 1671, 4512.
- Emami, A., Javanmardi, F., Akbari, A., Kojuri, J., Bakhtiari, H., Rezaei, T., ... & Falahati, F. (2021). Survival rate in hypertensive patients with COVID-19. *Clinical and experimental hypertension*, 43(1), 77-80.
- Etikan, I., Abubakar, S., & Alkassim, R. (2017). A review of life table construction. *Biom Biostat International Journal*, 5(3), 132.
- Freireich, E. J., Gehan, E., Frei lii, E. M. I. L., Schroeder, L. R., Wolman, I. J., ... & LEE, S. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. *Blood*, 21(6), 699-716.
- Gao, F., Manatunga, A. K., & Chen, S. (2001). A tree-structured method for analyzing multivariate survival data. In *Proceedings of the Annual Meeting of the American Statistical Association*.

- García, J. (2012). Análisis de supervivencia aplicado al estudio de la mortalidad en injertos de inchi (*Caryodendron orinocense* Karsten). *Revista Científica UDO Agrícola*, 12(4), 759-769.
- Greenwood, M. (1926). The natural duration of cancer (report on public health and medical subjects no 33). London: Stationery Office.
- Griffin, W. A., & Gardner, W. (1989). Analysis of behavioral durations in observational studies of social interaction. *Psychological bulletin*, 106(3), 497.
- Hougaard, P. (2012). Analysis of multivariate survival data. *Springer Science & Business Media*.
- Huang, Q. (2018). How to Improve the Survival Rate of Milk Calves. *Insight-Veterinary Science*, 1(1).
- Igwenagu, C. M. (2014). The Application of Life Table Functions: A Demographic Study. *IOSR Journal of Mathematics (IOSR-JM)*, 10 (1), 80-82.
- Julia, K. (2012). Survival analysis.. *Pediatrics in review / American Academy of Pediatrics*. 33. 172-4. 10.1542/pir.33-4-172.
- Kachman, S. D. (1999). Applications in survival analysis. *Journal of animal science*, 77(suppl\_2), 147-153.
- Kartsonaki, C. (2016). Survival analysis. *Diagnostic Histopathology*, 22(7), 263-270.
- Kaplan, E.L.; Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53, 457–481.
- Kleinbaum, D. G., & Klein, M. (2010). Survival análisis. A self-learning text. Springer, 3rd edition.
- Le, C. T. (1997). *Applied survival analysis*. John Wiley and sons. New York, United States of America
- Lee, E.T.; Wang, J. (2003) *Statistical Methods for Survival Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA; Volume 476.
- Lesem, M.D.; Tran-Johnson, T.K.; Riesenber, R.A.; Feifel, D.; Allen, M.H.; Fishman, R.; Spyker, D.A.; Kehne, J.H.; Cassella, J.V. (2011). Rapid acute treatment of agitation in individuals with schizophrenia: Multicentre, randomised, placebo-controlled study of inhaled loxapine. *Br. J. Psychiatry*. 198, 51–58.
- Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual review of public health*, 18(1), 83-104.
- Liedtke, C.; Mazouni, C.; Hess, K.R.; André, F.; Tordai, A.; Mejia, J.A.; Symmans, W.F.; Gonzalez-Angulo, A.M.; Hennessy, B.; Green, M.; *et al...*(2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol.* 26, 1275–1281.



- Massons, J. M. D. (1992). Una aplicación del análisis de la supervivencia en ciencias de la salud. *Anuario de psicología/The UB Journal of psychology*, (55), 109-142.
- Mitra, A., Pal, N., Saha, A., Chaudhuri, T. R., Fazli, P., & Zaman, S. (2017). Survival rate of mangroves: A proxy to assess ecosystem health.
- Moreira, D. R., Uberti, A. C. M. G., Gomes, A. R. Q., Ferreira, M. E. S., da Silva Barbosa, A., Varela, E. L. P., ... & Percário, S. (2021). Dexamethasone increased the survival rate in Plasmodium berghei-infected mice. *Scientific reports*, 11(1), 1-12.
- Muriel de la Riva, S., Cantalapiedra, M., & Lopéz, F. (2010). *Towards Advanced Methods for Computing Life Tables*. Technical Report.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14, 945-966.
- Pavía, J. M., Morillas, F., & Lledó, J. (2012). Introducing migratory flows in life table construction. *SORT*, 36(1), 103-114.
- Praxedes, Q & Reis, Y & Junior, A & Eduardo-Santos, Anderson. (2019). Survival rates of arboreal ants in an urbanized area.
- Prinja, S., Gupta, N., & Verma, R. (2010). Censoring in clinical trials: review of survival analysis techniques. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 35(2), 217.
- Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11), 2389-2430.
- Rivas López, M., & López Fidalgo, J. (2000). *Análisis de supervivencia*. La Muralla, 1ª edición.
- Robinson, B. H., Phillips, L. M., & Powell, A. N. (2019). Energy intake rate influences survival of Black Oystercatcher *Haematopus bachmani* broods. *Marine Ornithology*, 47, 277-283.
- Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *psychological Bulletin*, 110(2), 268.
- Vjollca, I., Moltine, P & Marena, T. (2019). Aegilops Seeds Germinability And Their Seedlings Drought Survival Rate. International Conference of Ecosystems (ICE2019)
- Yazdanpanah-Goharrizi, L., Zorriehzaha, M. J., & Sheikh-Asadi, M. (2020). The challenge of *Aeromonas hydrophila* and bacteriophage AH $\phi$ 3 to Rainbow trout (*Oncorhynchus mykiss*) in order to evaluate of survival rate. *International Journal of Infectious Diseases*, 101, 5.

## ANEXO TABLAS

**Tabla de supervivencia**

	Hora	Estado	Proporción acumulada que sobrevive en el tiempo		N de eventos acumulados	N de casos restantes
			Estimación	Desv. Error		
1	1,000	RECAIDA	.	.	1	41
2	1,000	RECAIDA	,952	,033	2	40
3	2,000	RECAIDA	.	.	3	39
4	2,000	RECAIDA	,905	,045	4	38
5	3,000	RECAIDA	,881	,050	5	37
6	4,000	RECAIDA	.	.	6	36
7	4,000	RECAIDA	,833	,058	7	35
8	5,000	RECAIDA	.	.	8	34
9	5,000	RECAIDA	,786	,063	9	33
10	6,000	RECAIDA	.	.	10	32
11	6,000	RECAIDA	.	.	11	31
12	6,000	RECAIDA	,714	,070	12	30
13	6,000	CENSURADO	.	.	12	29
14	7,000	RECAIDA	,690	,072	13	28
15	8,000	RECAIDA	.	.	14	27
16	8,000	RECAIDA	.	.	15	26
17	8,000	RECAIDA	.	.	16	25
18	8,000	RECAIDA	,591	,076	17	24
19	9,000	CENSURADO	.	.	17	23
20	10,000	RECAIDA	,565	,077	18	22
21	10,000	CENSURADO	.	.	18	21
22	11,000	RECAIDA	.	.	19	20
23	11,000	RECAIDA	,512	,079	20	19
24	11,000	CENSURADO	.	.	20	18
25	12,000	RECAIDA	.	.	21	17
26	12,000	RECAIDA	,455	,080	22	16
27	13,000	RECAIDA	,426	,080	23	15
28	15,000	RECAIDA	,398	,079	24	14
29	16,000	RECAIDA	,369	,078	25	13
30	17,000	RECAIDA	,341	,077	26	12
31	17,000	CENSURADO	.	.	26	11
32	19,000	CENSURADO	.	.	26	10
33	20,000	CENSURADO	.	.	26	9
34	22,000	RECAIDA	.	.	27	8
35	22,000	RECAIDA	,265	,077	28	7
36	23,000	RECAIDA	.	.	29	6
37	23,000	RECAIDA	,189	,071	30	5
38	25,000	CENSURADO	.	.	30	4
39	32,000	CENSURADO	.	.	30	3
40	32,000	CENSURADO	.	.	30	2
41	34,000	CENSURADO	.	.	30	1
42	35,000	CENSURADO	.	.	30	0

Tabla 1: Tabla de supervivencia de pacientes con leucemia obtenida mediante el método de Kaplan-Meier con el programa estadístico SPSS a partir de los datos recopilados por Freireich *et al.* (1963).

**Tabla de supervivencia**

GRUPO	Hora	Estado	Proporción acumulada que sobrevive en el tiempo		N de eventos acumulados	N de casos restantes	
			Estimación	Desv. Error			
TRATAMIENTO	1	6,000	RECAIDA	.	.	1	20
	2	6,000	RECAIDA	.	.	2	19
	3	6,000	RECAIDA	,857	,076	3	18
	4	6,000	CENSURADO	.	.	3	17
	5	7,000	RECAIDA	,807	,087	4	16
	6	9,000	CENSURADO	.	.	4	15
	7	10,000	RECAIDA	,753	,096	5	14
	8	10,000	CENSURADO	.	.	5	13
	9	11,000	CENSURADO	.	.	5	12
	10	13,000	RECAIDA	,690	,107	6	11
	11	16,000	RECAIDA	,627	,114	7	10
	12	17,000	CENSURADO	.	.	7	9
	13	19,000	CENSURADO	.	.	7	8
	14	20,000	CENSURADO	.	.	7	7
	15	22,000	RECAIDA	,538	,128	8	6
	16	23,000	RECAIDA	,448	,135	9	5
	17	25,000	CENSURADO	.	.	9	4
	18	32,000	CENSURADO	.	.	9	3
	19	32,000	CENSURADO	.	.	9	2
	20	34,000	CENSURADO	.	.	9	1
	21	35,000	CENSURADO	.	.	9	0
PLACEBO	1	1,000	RECAIDA	.	.	1	20
	2	1,000	RECAIDA	,905	,064	2	19
	3	2,000	RECAIDA	.	.	3	18
	4	2,000	RECAIDA	,810	,086	4	17
	5	3,000	RECAIDA	,762	,093	5	16
	6	4,000	RECAIDA	.	.	6	15
	7	4,000	RECAIDA	,667	,103	7	14
	8	5,000	RECAIDA	.	.	8	13
	9	5,000	RECAIDA	,571	,108	9	12
	10	8,000	RECAIDA	.	.	10	11
	11	8,000	RECAIDA	.	.	11	10
	12	8,000	RECAIDA	.	.	12	9
	13	8,000	RECAIDA	,381	,106	13	8
	14	11,000	RECAIDA	.	.	14	7
	15	11,000	RECAIDA	,286	,099	15	6
	16	12,000	RECAIDA	.	.	16	5
	17	12,000	RECAIDA	,190	,086	17	4
	18	15,000	RECAIDA	,143	,076	18	3
	19	17,000	RECAIDA	,095	,064	19	2
	20	22,000	RECAIDA	,048	,046	20	1
	21	23,000	RECAIDA	,000	,000	21	0

Tabla 2: Tabla de supervivencia en la que se comparan los dos grupos de pacientes, obtenida mediante el método de Kaplan-Meier con el programa estadístico SPSS a partir de los datos recopilados por Freireich *et al.* (1963).