

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA
DEPARTAMENTO DE ESTADÍSTICA.



TRABAJO DE GRADUACIÓN:

*“Análisis de Supervivencia y su Aplicación
para Predecir la Calidad de Vida de los
Nacidos Extremadamente Prematuros”*

PRESENTADO POR:

Estalin Ademir Mejía Hernández

PARA OPTAR AL TÍTULO:

LICENCIADO EN ESTADÍSTICA.

CIUDAD UNIVERSITARIA, SAN SALVADOR, EL SALVADOR, MARZO 2009.

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA
DEPARTAMENTO DE ESTADÍSTICA.



TRABAJO DE GRADUACIÓN:

*“Análisis de Supervivencia y su Aplicación
para Predecir la Calidad de Vida de los
Nacidos Extremadamente Prematuros”*

PRESENTADO POR:

Estalin Ademir Mejía Hernández

PARA OPTAR AL TÍTULO:

LICENCIADO EN ESTADÍSTICA.

ASESOR:

DR. JOSÉ NERYS FUNES TORRES

CIUDAD UNIVERSITARIA, SAN SALVADOR, EL SALVADOR, MARZO 2009.

AUTORIDADES DE LA UNIVERSIDAD

RECTOR : ING. RUFINO ANTONIO QUEZADA.

VICE - RECTOR ACADÉMICO : MSC. MIGUEL ÁNGEL PÉREZ.

FISCAL GENERAL : DR. RENE MADECADEL PERLA.

AUTORIDADES DE LA FACULTAD

DECANO : DR. RAFAEL GÓMEZ ESCOTO.

SECRETARIA : LICDA. MARIA TRINIDAD TRIGUEROS DE CASTRO.

DIRECTOR DE ESCUELA : ING. CARLOS MAURICIO CANJURA

AGRADECIMIENTOS

A Dios por haberme brindado todo para culminar este esfuerzo, en especial el regalo de mi hija: *KRISTHELL BELEN* quien se convierte en el último aliciente para culminar este proyecto y a la Virgen Maria por ser la intercesora de mis peticiones y las oraciones de mi familia.

A mis padres: *MAURICIO MEJÍA LÓPEZ Y HORTENSIA HERNÁNDEZ DE MEJÍA* ya que sin su apoyo y ánimos no hubiese logrado culminar este proyecto de vida.

A mis hermanos: *JOSÉ MAURICIO MEJÍA HERNÁNDEZ Y MARIA IDALIA MEJÍA HERNÁNDEZ* por tener la confianza en mí.

A mi asesor: *Dr. JOSÉ NERYS FUNES TORRES* por brindarme parte de sus conocimientos, además por motivarme a continuar el proyecto aunque las adversidades se presentaran durante su realización.

A mis Maestros, Compañeros, Amigos y todas aquellas personas que forman o han formado parte de mi vida.

A la Asociación Nacional de Árbitros de Fútbol y al Arbitraje Federado ya que se convirtieron en el pilar económico necesario durante gran parte de mis estudios, así como la fuente de grandes amistades y de personas con sabios y oportunos consejos.

DEDICADO A

† Mis abuelas: *Ángela Mejía y Victorina Hernández* por haberme educado durante mi niñez y adolescencia.

† Al don *Álvaro Magaña y su Familia* ya que fueron los primeros maestros que tuve y que cultivaron en mí el deseo de Superación y la entrega a los estudios.

Mi hija, para que entienda que lo bueno siempre conlleva un esfuerzo, pero al final, la gratitud por los consejos y castigos que los padres brindan se agradecen eternamente.

Mi Sobrino: *MICHAEL RIQUELMER*, ya que empieza su vida académica y que esto sirva como motivación para la culminación de sus estudios.

ÍNDICE

CAPITULO I: INTRODUCCIÓN AL ANÁLISIS DE SUPERVIVENCIA.	1
1.1 INTRODUCCIÓN	1
1.2 ASPECTOS BÁSICOS DEL ANÁLISIS DE SUPERVIVENCIA	3
1.2.1 OBTENCIÓN DE LOS DATOS	3
1.2.2 TIPOS DE DATOS CENSURADOS. ESQUEMAS DE CENSURA.....	4
1.3 CONCEPTOS PROBABILÍSTICOS BÁSICOS DE LA SUPERVIVENCIA.	8
1.3.1 MODELOS CONTINUOS.....	8
1.3.2 MODELOS DISCRETOS	11
1.4 INTERPRETACIÓN DE LA FUNCIÓN DE RIESGO	13
1.5 ALGUNAS DISTRIBUCIONES DE PROBABILIDAD BÁSICAS.	15
1.5.1 DISTRIBUCIÓN EXPONENCIAL.	16
1.5.2 DISTRIBUCIÓN DE WEIBULL.....	20
1.5.3 DISTRIBUCIÓN LOGNORMAL.....	23
1.6 ESTIMACIÓN PUNTUAL Y POR INTERVALOS PARA LOS PARÁMETROS DE FUNCIONES DE SUPERVIVENCIA.	27
1.7 MODELOS PARA PRUEBAS ACELERADAS	35
CAPITULO II: ESTIMACIÓN NO PARAMÉTRICA DE LA SUPERVIVENCIA	42
2.1 INTRODUCCIÓN	42
2.2 TABLAS DE VIDA	44
2.3 TABLAS DE VIDA PARA VARIOS GRUPOS	54
2.4 ESTIMACIÓN DE LAS FUNCIONES DE SUPERVIVENCIA	68
2.4.1 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA: MÉTODO ACTUARIAL	75
2.4.2 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA: PRODUCTO LÍMITE DE KAPLAN Y MEIER.	83
2.5 PRECISIÓN DE LAS ESTIMACIONES	86
2.6 ESTIMACIÓN DE OTRAS FUNCIONES Y PARÁMETROS DE INTERÉS	92
2.7 INFERENCIA SOBRE LA FUNCIÓN DE SUPERVIVENCIA.	96
2.7.1 ANÁLISIS GRÁFICO DE LA ADECUACIÓN DE UNA DISTRIBUCIÓN.	96
2.7.2 TEST DE LOG RANK.....	97
2.7.3 RIESGO RELATIVO	103

CAPITULO III: APLICACIÓN DEL ANÁLISIS DE SUPERVIVENCIA.....	105
3.1 INTRODUCCIÓN	105
3.2 DESCRICIÓN DE LA BASE DE DATOS Y TERMINOLOGÍA MÉDICA.....	107
3.2.1 NOMBRE DE LAS VARIABLES EN LA BASE DE DATOS.	107
3.2.2 PRINCIPALES DEFINICIONES MÉDICAS UTILIZADAS.....	109
3.3 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES.....	111
3.4 TABLAS DE CONTINGENCIA DE ALGUNAS VARIABLES	122
3.5 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA.....	129
3.5.1 TABLAS DE VIDA Y GRÁFICOS DE LAS FUNCIONES DE SUPERVIVENCIA Y DE RIESGO	129
3.5.2 MÉTODO DE KAPLAN – MEIER, GRÁFICOS DE SUPERVIVENCIA Y DE RIESGO.	134
3.5.3 ESQUEMA DE CENSURA	139
3.6 CONCLUSIONES	140
ANEXOS.....	141

CAPITULO I: INTRODUCCIÓN AL ANÁLISIS DE SUPERVIVENCIA.

1.1 INTRODUCCIÓN

El origen del análisis de supervivencia se remonta al siglo XVII con la construcción de las Tablas de Vida, y se debe primordialmente a que en muchas aplicaciones el suceso de interés era entonces la muerte. No obstante el Análisis de Supervivencia que actualmente estudiamos se desarrolló inicialmente mediante aproximaciones semiparamétricas (Cox, 1972; Prentice y Gloeckler, 1978), sin embargo los modelos paramétricos (Ducrocq et al., 1998a,b) presentan ventajas importantes en cuanto a tiempo de computación (Ducrocq et al., 2000).

Las técnicas de Análisis de Supervivencia se fundamentan en dos distribuciones de probabilidad específica, siendo estas: las funciones de Supervivencia y de Riesgo. La función de Supervivencia en el momento t , $S(t)$, se define como la probabilidad de que un individuo sobreviva más de t unidades de tiempo (Cox, 1972). Por otro lado, la función de Riesgo $h(t)$ representa la tasa instantánea de muerte en el momento t condicionada a que el individuo a sobrevivido más de t ($T > t$).

El rasgo específico del análisis estadístico en el campo de la medicina, por ejemplo es la necesidad de realizar inferencia a partir de muestras en las que las observaciones de la variable, aparecen observaciones incompletas, parciales o censuradas. La obtención de muestras completas suele requerir ensayos demasiado largos, por lo que es habitual terminar los experimentos cuando se ha observado un determinado porcentaje de “fallos” o diseñarlos con un horizonte temporal prefijado, de tal forma que frecuentemente no se observe el “fallo” de un número importante de elementos de la muestra.

En los ensayos clínicos la censura aparece debido al diseño de los mismos; en estos estudios los pacientes suelen incorporarse en diferentes instantes y probablemente no se observa el fallo de aquellos elementos o individuos que se han incorporado más tarde. Además, es frecuente que durante el desarrollo de la prueba algunos individuos abandonen el estudio por lo que tampoco es posible conocer su tiempo de fallo exacto, sino, que éste no se había observado hasta el momento del abandono.

En este capítulo realizaremos una introducción a la teoría necesaria del Análisis de Supervivencia, presentando las definiciones y el comportamiento gráfico de la función de Riesgo y de Supervivencia, seguidamente mostraremos las distribuciones principales que existen en el Análisis de Supervivencia, entre ellas: Exponencial, Weibull y Log Normal, para finalmente desarrollar la teoría sobre la Estimación Puntual y por Intervalos para los Parámetros de Funciones de Supervivencia y los Ajustes de Funciones de Supervivencia y Test de Bondad de Ajuste.

Por otra parte, se describirá brevemente el proceso que se realizó para crear la Base de Datos que ocuparemos para aplicar toda la teoría del Análisis de Supervivencia, además daremos la base teórica que servirá de forma inductiva para el Capítulo II.

1.2 ASPECTOS BÁSICOS DEL ANÁLISIS DE SUPERVIVENCIA.

En este apartado daremos a conocer el proceso que se siguió para la creación de la base de datos que se utilizará para realizar la aplicación del Análisis de Supervivencia, además comentaremos sobre los tipos de censura que se tienen en este tipo de estudio, mostrando esquemas en los que se explicará qué es una censura y qué papel juegan en el Análisis de Supervivencia. Por otra parte ejemplarizaremos algunos casos sencillos que servirán como guía para familiarizarnos con algunos conceptos básicos en nuestro estudio.

1.2.1 OBTENCIÓN DE LOS DATOS

Los datos de este estudio se han obtenido mediante la revisión de todos los expedientes clínicos de aquellas pacientes que tuvieron un parto dentro del Hospital Nacional de Maternidad desde el año 1999 hasta el año 2002, ya sea por cesárea o vaginal. La única restricción para ser tomado en cuenta dentro de este estudio fue que el egreso del recién nacido fuese vivo, además que al nacer su peso estuviera dentro del rango de 500 a 1000 gramos, a fin de tener una población bajo estudio que clínicamente se conoce como un recién nacido **extremadamente prematuro**.

La base de datos se forma mediante la información obtenida en una hoja construida por un equipo médico del Departamento de Neonatología y reúne según éstos las variables más importantes desde el nacimiento hasta el egreso del recién nacido, en esta hoja se presentan tres etapas del recién nacido, siendo estas: Etapa del nacimiento, en la cual se toman los primeros síntomas de vida, tratamiento a posibles complicaciones desarrolladas por la prematuridad y una última parte que consiste en conocer el estado de salud con el cual el recién nacido egresa del Hospital.

Dentro de este estudio no se ha tomado en cuenta el origen geográfico de la madre ni la edad gestacional del embarazo, y como censuras se tienen tres tipos: niños que al ser buscados ya han fallecido, niños que no pudieron ser localizados o no quisieron participar de este estudio por consideraciones particulares de la madre y aquellos que al finalizar este estudio aun se encuentran vivos. Debe de tomarse en cuenta que como en todo estudio

clínico, debido a la dificultad para disponer al comienzo del estudio de un número suficiente de pacientes que satisfagan todos los requisitos, la incorporación de estos se produce de forma escalonada, en distintos instantes de tiempo.

Es por ello, que en este caso su incorporación se da en el momento de su nacimiento, como se podrá notar la edad que tendrá cada niño al finalizar este estudio será diferente y por lo tanto, se puede esperar que su estado físico y de salud sea diferente, aun más si tomamos en cuenta los posibles cuidados o control clínico que las madres han tenido con ellos.

1.2.2 TIPOS DE DATOS CENSURADOS. ESQUEMAS DE CENSURA

Los datos correspondientes a estudios de Análisis de Supervivencia presentan una particularidad que dificulta su análisis estadístico con los métodos clásicos de análisis exploratorio o estadística inferencial. Esta peculiaridad es la presencia de datos censurados: solo se conoce el tiempo de fallo para una fracción, que puede ser pequeña, de los individuos de la muestra, mientras que del resto se dispone solo de información parcial, habitualmente que el “tiempo de vida” es mayor que un valor dado.

Una observación se dice **censurada a la derecha de L** , si se desconoce el valor exacto de la observación y solo se sabe que ésta es mayor que L . Análogamente, una observación se dice **censurada a la izquierda de L** , si solo se sabe que la observación es menor que el valor L . La censura a la derecha es mucho más frecuente que la censura a la izquierda.

En algunos experimentos, dependiendo del tipo de problema y el tipo de seguimiento, aparecen datos censurados en un intervalo t_I, t_D ; es decir, que solo se sabe que el suceso de interés ocurrió en el intervalo $t_I < T < t_D$.

En resumen los tipos de Censura se pueden clasificar en dos esquemas, los cuales son básicos para establecer estos límites.

Censura de tipo I. En este esquema el experimento se programa con una duración, C , establecida a priori. El tiempo de fallo de un individuo se observará, si es menor o igual que ese valor prefijado. En otro caso, la observación correspondiente será censurada, con valor C , y la denotaremos C^* .

Censura de tipo II. En los ensayos realizados bajo un esquema de tipo II, con “ n ” componentes idénticos, el ensayo finaliza en el momento en que se produce el r -ésimo fallo $1 \leq r \leq n$. Ese instante, t_r será el valor de los datos censurados correspondientes a los componentes que en ese momento sigan funcionando. De esta forma solo se conocen las r observaciones más pequeñas de la muestra y aparecen $n-r$ tiempos censurados en el valor t_r .

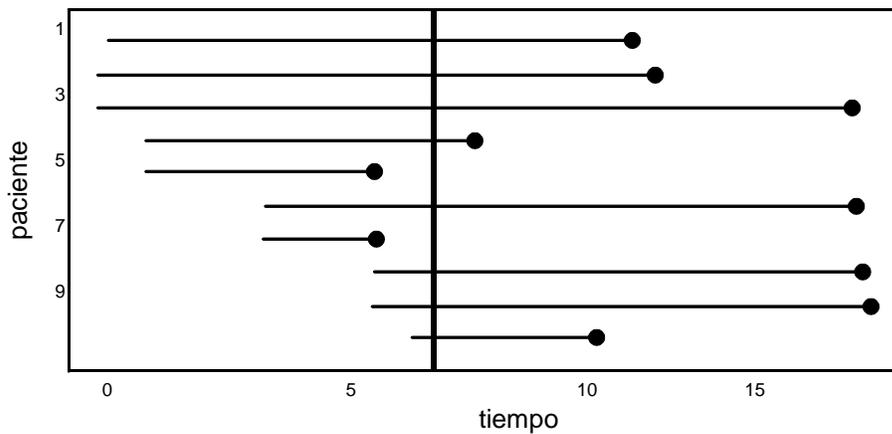
Es importante señalar que el valor de C en el esquema de tipo I y el valor de r (o la fracción r/n) que indica la tasa de censura en el esquema de tipo II deben fijarse antes de iniciar el experimento y no durante el transcurso del mismo dependiendo de los resultados que se observen. La necesidad de que el mecanismo de censura sea independiente de la observación del fenómeno, es un requisito imprescindible para la validez de las conclusiones.

Por otra parte, en los ensayos médicos aunque el experimento se diseñe con una limitación temporal como en el esquema I, es normal que los individuos se incorporen al ensayo en instantes aleatorios, cuando se dispone de los pacientes adecuados; además, es habitual que se produzcan abandonos durante la realización del ensayo. En consecuencia, las muestras resultantes son **múltiplemente censuradas**. Generalmente, este tipo de observaciones se presentan mediante un par de variables T, δ , donde T es el tiempo transcurrido desde la entrada del individuo al ensayo hasta su salida del mismo, δ es una variable binaria

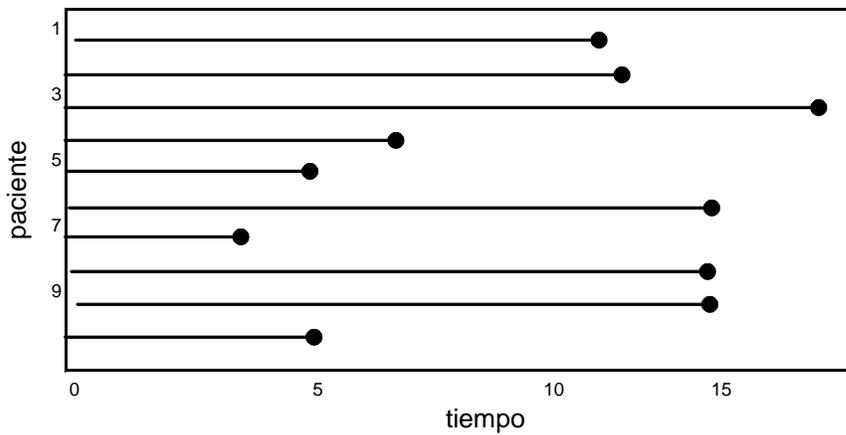
indicadora del tipo de observación, que toma el valor 1 si se ha observado el fallo y el valor 0 si se trata de una observación censurada.

FIGURA 1. ESQUEMA DE LOS TIEMPOS DE FALLO

a)



b)



Las figuras anteriores corresponden a un estudio de 18 meses de duración, pudiendo observar en la figura 1.a) que los dos pacientes que están tras la línea intermedia han abandonado el estudio, mientras que los cuatro que ya pasaron dicha línea pero aun no han llegado al final, representan pacientes que les ha ocurrido el evento de interés antes de finalizar el estudio, y los que observamos que la línea llega al final representan aquellos que aun permanecían vivos al final del estudio.

En el análisis de los tiempos de vida, el instante en el que se comenzó a medir cada observación no suele ser de interés, por lo que con frecuencia suelen representarse las observaciones con el mismo origen, como se muestra en la figura 1.b).

TABLA 1. DATOS CORRESPONDIENTES A FIGURA 1.

PACIENTE	TIEMPO DE ENTRADA	TIEMPO DE FALLO O CENSURA	ESTADO
1	0.0	11.8	F
2	0.0	12.5	C
3	0.4	18.0	C
4	1.2	6.6	F
5	1.2	4.4	C
6	3.0	18.0	C
7	3.4	4.9	F
8	4.7	18.0	C
9	5.0	18.0	C
10	5.8	10.1	F

En la tabla 1 se muestran los datos recogidos en un estudio de 18 meses de duración, en el que solo fueron admitidos pacientes durante los seis primeros meses, y que de forma análoga a lo que observamos en la Figura 1 podemos mencionar que el valor 18.0 corresponde a pacientes que al finalizar el estudio aun no les ha ocurrido el evento de interés; los valores 11.8, 12.5, 6.6 y 10.1 representan aquellos pacientes que ya les ha ocurrido el evento de interés; mientras que los valores de 4.4 y 4.9 representan aquellos pacientes que abandonaron el estudio. Los pacientes que ya les ocurrió el evento de interés representan el valor de Fracaso (F), mientras que los que han abandonado y los que aun no les a ocurrido el evento de interés representan valores Censurados (C) como lo podemos observar en la última columna de la tabla anterior.

1.3 CONCEPTOS PROBABILÍSTICOS BÁSICOS DE LA SUPERVIVENCIA.

1.3.1 MODELOS CONTINUOS.

Diremos que una variable aleatoria es continua si su espacio de resultados es un subconjunto continuo de \mathbb{R} . En las definiciones siguientes supondremos que el espacio de resultados de T es $\xi = [0; \infty)$.

La función $f(t)$, cuya gráfica es la curva límite que se obtiene para un número muy grande de observaciones y para una amplitud de intervalo muy pequeño, es la función de densidad de probabilidad para una variable aleatoria continua T , ya que la escala vertical se elige de tal manera que el área total bajo la curva es igual a uno. La función de densidad de probabilidad de una variable aleatoria continua T se define formalmente de la siguiente manera:

Definición 1. Función de Densidad

Si $f(t)$ es la función de densidad de probabilidad de la variable aleatoria continua T , entonces para cualquier $a, b \in \mathbb{R}$ se tiene:

$$1. \quad f(t) \geq 0, \quad -\infty < t < \infty, \quad (1)$$

$$2. \quad \int_{-\infty}^{\infty} f(t) dt = 1 \quad (2)$$

$$3. \quad P(a \leq T \leq b) = \int_a^b f(t) dt \quad (3)$$

Definición 2. Función de distribución.

Sea una variable aleatoria T de tipo continuo que toma un número infinito de valores sobre la recta real y cuya función de densidad es $f(t)$. Se define la **función de distribución acumulativa** de la variable aleatoria T , que notaremos por $F(t)$, como la probabilidad de que la variable aleatoria continua T tome valores menores o iguales a t , es decir:

$$F(t) = P(T \leq t) = \int_{-\infty}^t f(t)dt \tag{4}$$

Por lo tanto, la función de distribución acumulativa $F(t)$ es el área acotada por la función de densidad $f(t)$ que se encuentra a la izquierda de la recta $T = t$. la función $F(t)$ proporciona la probabilidad de que ocurra el evento de interés antes o en un momento t , del periodo de estudio que hemos fijado.

La distribución acumulativa $F(t)$, es una función no decreciente de los valores de la variable aleatoria con las siguientes propiedades.

1. $F(-\infty) = 0$; (5)

2. $F(\infty) = 1$; (6)

3. $P(a \leq T \leq b) = \int_a^b f(t)dt = F(b) - F(a)$; (7)

4. $dF(t)/dt = f(t)$. (8)

Que son fáciles de verificar, siendo la expresión (8) una consecuencia del teorema fundamental del cálculo integral.

Definición 3. La **función de Supervivencia $S(t)$** es la complementaria de la función de distribución, e indica la probabilidad de que un individuo sobreviva después de un tiempo t :

$$S(t) = P(T > t) = 1 - F(t) \tag{9}$$

Esta función es monótona decreciente y verifica que $S(0) = 1$ y $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$.

Definición 4. La **función de riesgo o función de tasa de fallo** $h(t)$, representa la evolución de la probabilidad de fallo en relación con la edad de los individuos, y se define como el cociente entre la función de densidad y la de supervivencia:

$$h(t) = \frac{f(t)}{S(t)} \quad (10)$$

Las expresiones siguientes, muestran las relaciones existentes entre las funciones más utilizadas,

$$\begin{aligned} f(t) &= \frac{dF(t)}{dt} = \frac{d(1-S(t))}{dt} \\ f(t) &= -\frac{d}{dt}S(t) = -S'(t) \end{aligned} \quad (11)$$

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} \\ h(t) &= -\frac{d}{dt} \ln S(t) \end{aligned} \quad (12)$$

Integrando la expresión (12) de 0 a t y usando el hecho $S(0) = 1$, tendremos:

$$-\int_0^t h(x)dx = \ln S(t) \quad (13)$$

ó

$$S(t) = \exp\left[-\int_0^t h(x)dx\right] \quad (14)$$

1.3.2 MODELOS DISCRETOS

En algunos casos, puede ser necesario tratar el tiempo de vida T como una variable aleatoria discreta; por ejemplo, cuando se mide el tiempo de vida de forma tal que el número de valores distintos que puede tomar la variable es pequeño y son frecuentes los empates en las observaciones. Supondremos que el espacio de resultados asociado es ahora $\xi = t_{(1)}, t_{(2)}, t_{(3)}, \dots$ donde $0 \leq t_{(1)} < t_{(2)} < \dots$

La distribución de T viene determinada por su función de probabilidad,

$$p_j = P(T = t_{(j)}) \quad j = 1, 2, \dots \quad (15)$$

con $p_j > 0$ y $\sum_j p_j = 1$

La función de supervivencia,

$$S(t) = P(T > t) = \sum_{j: t_{(j)} > t} p_j \quad (16)$$

A diferencia del caso continuo, el valor de la función de supervivencia difiere en los instantes de fallo del que se obtendrá; definiendo $S(t) = P(T \geq t)$.

La función de riesgo se define en los instantes $t_{(j)}$, como la probabilidad condicional de fallo en ese instante, dado que se ha llegado vivo a él:

$$h(t_{(j)}) = h_j = p(T = t_{(j)} / T \geq t_{(j)}) \quad j = 1, 2, \dots$$

Deduciendo, tenemos:

$$\begin{aligned}
 p(T = t_j / T \geq t_j) &= \frac{p(T = t_j \cap T \geq t_j)}{p(T \geq t_j)} \\
 &= \frac{p(T = t_j)}{p(T \geq t_j)} = \frac{P_j}{p(T > t_{j-1})} \\
 h(t_{(j)}) &= \frac{P_j}{S(t_{(j-1)})} \quad j = 1, 2, \dots \quad (17)
 \end{aligned}$$

además, $h(t) = 0$ si $t \neq t_{(j)}$

Como en el caso continuo, estas funciones determinan completamente la distribución de T .

Las relaciones existentes entre ellas son:

$$h_j = 1 - \frac{S(t_j)}{S(t_{(j-1)})} \quad j = 1, 2, \dots \quad (18)$$

$$S(t) = \prod_{j: t_{(j)} \leq t} (1 - h_j) \quad (19)$$

La expresión de función de riesgo acumulado es en este caso:

$$H(t) = \sum_{j: t_{(j)} \leq t} h_j \quad (20)$$

y no satisface la relación encontrada en el caso continuo pues, en general,

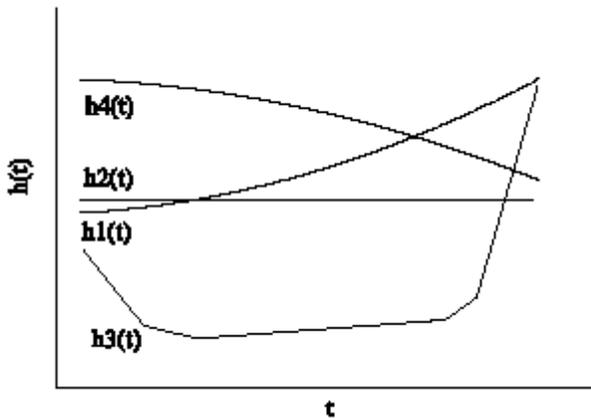
$$-\ln S(t) = - \sum_{j: t_{(j)} \leq t} \ln(1 - h_j) \neq \sum_{j: t_{(j)} \leq t} h_j \quad (21)$$

no obstante, como puede comprobarse desarrollando en serie $\ln(1-t)$, ambas expresiones producirán resultados próximos si los valores h_j son pequeños.

1.4 INTERPRETACIÓN DE LA FUNCIÓN DE RIESGO

Con el fin de tener una interpretación de la **función de riesgo** comentaremos cuatro ejemplos de poblaciones cuyo perfil de riesgo se corresponde con cada uno de los tipos de curvas que se muestran en la figura 2:

FIGURA 2. DISTINTAS FUNCIONES DE RIESGO



- $h_1(t)$: El conjunto de personas mayor de 65 años. Esta población presenta una función de riesgo creciente que nos indica que la tasa de fallo tiende a aumentar con el transcurso del tiempo. Por ejemplo, la probabilidad de que un individuo con 70 años viva más de 71, es mayor que la probabilidad de que un individuo con 80 viva más de 81.
- $h_2(t)$: Una población de individuos sanos entre los 20 y 40 años de edad, para los que el único riesgo de muerte, en la práctica, viene dado por distintos tipos de accidentes (laborales, deportivos, de tráfico, etc.). En esta población, la función de riesgo es prácticamente constante.

- $h3(t)$: Una población, por ejemplo la de los españoles nacidos en la década de los setenta, observada desde el nacimiento hasta la muerte. Esa población presentará una función de riesgo con forma de \cup , llamada también riesgo "bañera", típica de las tablas de vida poblacionales. Inicialmente se tiene un periodo con tasa de fallo alta, correspondiente a la etapa neonatal e infantil, que va decreciendo hasta estabilizarse. El riesgo permanece bajo y aproximadamente constante, hasta una cierta edad, en torno a los 40 años, a partir de la cual comienza a aumentar con el tiempo.
- $h4(t)$: Una población de personas jóvenes que padece cierto defecto congénito y que es sometida a un proceso quirúrgico complicado para corregirlo, analizada mientras dura el periodo de recuperación. Esta población presentará una tasa de riesgo decreciente ya que en estos casos, el principal riesgo de muerte aparece como consecuencia de la intervención o de sus complicaciones inmediatas.

1.5 ALGUNAS DISTRIBUCIONES DE PROBABILIDAD BÁSICAS.

En este apartado se revisan algunas de las distribuciones de probabilidad más empleadas en Análisis de Supervivencia. En principio, cualquier distribución no negativa se puede utilizar para modelizar el tiempo de vida de cierto suceso de interés; el objetivo es disponer de un conjunto de distribuciones lo suficientemente flexibles para adaptarse a los distintos tipos de datos y lo más sencillas posibles para facilitar su análisis.

En general, la distribución más utilizada en Estadística es la distribución Normal. En el Análisis de Supervivencia la distribución de referencia es la Exponencial. Las buenas propiedades de esta distribución, como consecuencia de su “ausencia de memoria”, permiten simplificar los problemas de inferencia; por el mismo motivo su aplicación práctica es limitada, siendo más utilizadas generalizaciones suyas como las distribuciones Weibull o Gamma.

Presentamos las distribuciones bajo la hipótesis de que el rango de valores de T es $\xi = [0, \infty)$. Todas las distribuciones de tiempo de fallo admiten una versión más general en la que aparece un nuevo parámetro G , llamado umbral o tiempo de garantía, que puede tomar cualquier valor no negativo. Esta versión generalizada, en la que el rango de T es $[G; \infty)$, se obtiene al considerar que $T - G$ tiene una distribución con rango $[0; \infty)$.

Los modelos con parámetro de garantía tienen interés en situaciones en las que el conocimiento previo del fenómeno permite suponer que el riesgo de fallo en el intervalo $[0; G)$ es nulo. Generalmente el valor de G es desconocido y es necesario estimarlo junto con los demás parámetros.

Si se encuentra una distribución de probabilidad que representa bien los datos, se pueden aplicar métodos de inferencia paramétricos basados en dicha distribución. Este tipo de análisis es más frecuente en el campo de la Fiabilidad que en el Análisis de Supervivencia.

1.5.1 DISTRIBUCIÓN EXPONENCIAL.

La distribución más importante, en los estudios de supervivencia, es la distribución exponencial. En Análisis de Supervivencia, juega un papel análogo al de la distribución normal en otras áreas de la estadística.

Definición 5. Diremos que una variable aleatoria T , de tipo continuo, sigue una **distribución exponencial** de parámetro λ , siendo $\lambda \in R$ y $\lambda > 0$, si su función de densidad es:

$$f(t) = \lambda e^{-\lambda t} \quad \text{para } t \geq 0, \lambda > 0 \quad (22)$$

Esta distribución es un caso particular de la distribución Gamma, $\Gamma(p, \lambda)$ para $p = 1$, es decir, que corresponde a una $\Gamma(1, \lambda)$. Hecho que tendremos en cuenta para el estudio de la media y la varianza.

La distribución Exponencial está bastante relacionada con la de Poisson, así pues si el número de sucesos que ocurren en un determinado intervalo sigue una distribución de Poisson, entonces la variable aleatoria que representa el tiempo entre ocurrencia de sucesos sigue una distribución exponencial.

Evidentemente, la función de densidad, está bien definida, pues para $t > 0, f(t) > 0$ y además

$$\int_0^{\infty} \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_0^{\infty} = 1 \quad (23)$$

La correspondiente **función de distribución** $F(t)$, viene expresada por:

$$F(t) = P(T \leq t) = \int_0^t \lambda e^{-\lambda t} dt \quad t \geq 0 \quad (24)$$

y ocupando el razonamiento de la expresión (23), tendremos:

$$F(t) = 1 - e^{-\lambda t} \quad t \geq 0 \quad (25)$$

La **función de Supervivencia**, que es la probabilidad de sobrevivir a un instante t , la obtendremos partiendo de la expresión (9) para tener que:

$$S(t) = 1 - F(t) \quad (26)$$

sustituyendo la expresión (25) en (26), tendremos:

$$S(t) = e^{-\lambda t} \quad (27)$$

La **función de riesgo**, que da la tasa condicional de fallo o muerte de un elemento que no haya fallado antes del instante t , viene dado por la expresión:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \quad t \geq 0 \quad (28)$$

es decir:

$$h(t) = \lambda \quad t \geq 0 \quad (29)$$

Lo que nos muestra que el riesgo es constante a lo largo del tiempo. A esta propiedad de la distribución exponencial es a lo que se le suele llamar **pérdida de memoria**.

Para llegar a obtener la media calcularemos previamente el momento de orden r respecto al origen, utilizaremos la función de densidad de la **función gamma** y algunas propiedades de esta función.

Definición 6. Diremos que una variable T , de tipo continuo, sigue una **distribución gamma** de parámetros p (*parámetro de forma*) y λ (*parámetro de escala*), siendo $p, \lambda \in \mathbb{R}$ y $p > 0$ y $\lambda > 0$, si su función de densidad es:

$$f(t) = \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t} \quad \text{donde}$$

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad p > 0 \quad (30)$$

La función $\Gamma(p)$ es convergente para $p > 0$, siendo p un número real positivo pero no necesariamente entero. Algunas propiedades de esta función son las siguientes:

1. $\Gamma(p) = (p-1)!$ si n es un entero positivo

2. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

3. $\Gamma(1) = 1$

4. $\Gamma(p+1) = p\Gamma(p)$

entonces, el momento de orden r con respecto al origen está dada por:

$$\begin{aligned} E\left[T^r\right] &= \int_0^{\infty} t^r f(t) dt = \int_0^{\infty} t^r \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t} dt \\ &= \frac{\lambda^p}{\Gamma(p)} \int_0^{\infty} t^{p+r-1} e^{-\lambda t} dt = \frac{\lambda^p}{\Gamma(p)} \cdot \frac{\Gamma(p+r)}{\lambda^{p+r}} \\ &= \frac{\Gamma(p+r)}{\lambda^r \Gamma(p)} \end{aligned} \quad (31)$$

Expresión que nos permite calcular fácilmente la *media*, solamente hacemos $r = 1$

$$E T = \frac{\Gamma(p+1)}{\lambda\Gamma(p)} = \frac{p}{\lambda} \quad \text{por propiedades de función gamma} \quad (32)$$

Para obtener la varianza obtendremos previamente el momento de orden dos, haciendo para ello $r = 2$ en la expresión (31)

$$E[T^2] = \frac{\Gamma(p+2)}{\lambda^2\Gamma(p)} \quad \text{por propiedades de función gamma}$$

$$E[T^2] = \frac{(p+1)p}{\lambda^2} \quad (33)$$

luego la varianza será:

$$\begin{aligned} \text{Var}(T) &= E[T^2] - E T^2 \\ &= \frac{(p+1)p}{\lambda^2} - \frac{p^2}{\lambda^2} \\ &= \frac{p}{\lambda^2} \end{aligned} \quad (34)$$

Entonces, si hacemos $p = 1$ en las expresiones (32) y (34) obtendremos que la media y la varianza de una variable aleatoria con distribución exponencial, están dadas por:

$$\begin{aligned} E T &= \frac{1}{\lambda} \\ \text{Var } T &= \frac{1}{\lambda^2} \end{aligned} \quad (35)$$

Los valores de la Función Gamma se encuentran tabulados en varios libros de estadística.

1.5.2 DISTRIBUCIÓN DE WEIBULL.

La distribución de Weibull fue establecida por el físico suizo del mismo nombre, quien demostró, con base en una evidencia empírica, que el esfuerzo al que se someten los materiales puede modelarse de manera adecuada mediante el empleo de esta distribución.

Definición 7. Diremos que una variable aleatoria T sigue la distribución de Weibull si su función de densidad $f(t)$ tiene la forma:

$$f(t) = \frac{\alpha}{\theta^\alpha} t^{\alpha-1} e^{-t/\theta} \quad \text{si } t > 0 \quad \alpha, \theta > 0 \quad (36)$$

En la expresión anterior al parámetro α se le llama **parámetro de forma**, en tanto que θ es el **parámetro de escala**.

La Función de Distribución de Weibull

$$F(x) = \frac{\alpha}{\theta^\alpha} \int_0^x t^{\alpha-1} e^{-t/\theta} dt \quad (37)$$

puede obtenerse en forma cerrada mediante la evaluación directa de la integral en (37), esto es:

$$\begin{aligned} F(t) &= \left[\frac{\alpha}{\theta^\alpha} \left(-\frac{\theta^\alpha}{\alpha} \right) e^{-x/\theta} \right]_0^t \\ &= 1 - e^{-t/\theta} \quad t \geq 0 \end{aligned} \quad (38)$$

Luego la función de Supervivencia viene dada por:

$$S(t) = P(T > t) = 1 - F(t) = e^{-t/\theta^\alpha} \quad t \geq 0 \quad (38)$$

Por lo tanto, la **función de riesgo h(t)**, que da la probabilidad de sobrevivir a un instante t , es:

$$h(t) = \frac{\frac{\alpha}{\theta^\alpha} t^{\alpha-1} e^{-t/\theta^\alpha}}{e^{-t/\theta^\alpha}}$$

$$h(t) = \frac{\alpha}{\theta^\alpha} t^{\alpha-1} \quad t > 0 \quad (39)$$

para determinar el valor de la media y la varianza, encontraremos primero el r -ésimo momento central alrededor del origen.

$$\begin{aligned} \mu_r &= E[T^r] = \int_0^\infty t^r f(t) dt \\ &= \frac{\alpha}{\theta^\alpha} \int_0^\infty t^{\alpha+r-1} e^{-t/\theta^\alpha} dt \end{aligned} \quad (40)$$

En (40), sea $u = t/\theta^\alpha$; entonces $t = \theta u^{1/\alpha}$ y $dt = \theta/\alpha u^{1/\alpha-1} du$.

El resultado es:

$$\begin{aligned}\mu_r &= \frac{\alpha}{\theta^\alpha} \int_0^\infty \theta u^{\frac{1}{\alpha} \alpha + r - 1} e^{(-u)} \frac{\theta}{\alpha} u^{\frac{1}{\alpha} - 1} du \\ &= \theta^r \int_0^\infty u^{r/\alpha} e^{(-u)} du \\ &= \theta^r \Gamma\left(1 + \frac{r}{\alpha}\right)\end{aligned}\tag{41}$$

Entonces la media de T , es:

$$E T = \theta \Gamma\left(1 + \frac{1}{\alpha}\right)\tag{42}$$

y la varianza de T , es el resultado de evaluar:

$$Var T = \theta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]\tag{43}$$

1.5.3 DISTRIBUCIÓN LOGNORMAL.

La distribución Log-Normal está estrechamente relacionada con la normal de tal manera que la función de densidad de la distribución Log-Normal coincide con la función de densidad de una variable aleatoria cuyo logaritmo sigue una distribución normal.

Definición 8. Diremos que una variable aleatoria T , de tipo continua, no negativa, sigue una **distribución Log-Normal** de parámetros μ y σ , si la variable aleatoria $Y = \ln T$ sigue una distribución $N(\mu, \sigma)$, con

$$-\infty < \mu < \infty \text{ y } \sigma > 0$$

Entonces la función de densidad se obtiene a partir de la propia definición, teniendo que:

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\ln T \leq \ln t) \\ &= P(Y \leq \ln t) = F_Y(\ln t) \end{aligned} \quad t > 0 \quad (44)$$

siendo nula para valores $t \leq 0$.

Ahora, derivando la expresión anterior y sustituyendo se tiene:

$$f_T(t) = \frac{1}{t} f_Y(\ln t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, t > 0 \quad (45)$$

que es la **función de densidad** de la distribución Log-Normal de parámetros μ y σ o función de Cobb- Douglas.

Si hacemos $a = e^{-\mu}$ entonces $-\mu = \ln a$, y la ecuación (45) se transforma en:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln at)^2}{2\sigma^2}} \quad (46)$$

La **función de distribución** se obtiene a partir de la expresión (44), pues tipificando y teniendo en cuenta que la variable aleatoria $\frac{Y-\mu}{\sigma} = Z$, es $N(0,1)$, se tiene:

$$\begin{aligned} F_T(t) &= F_Y(\ln t) = P(Y \leq \ln t) \\ &= P\left(\frac{Y-\mu}{\sigma} \leq \frac{\ln t - \mu}{\sigma}\right) = P\left(Z \leq \frac{\ln t - \mu}{\sigma}\right) \\ F_Y(t) &= F\left(\frac{\ln t - \mu}{\sigma}\right) \equiv F\left(\frac{\ln at}{\sigma}\right) \end{aligned} \quad (47)$$

Lo cual nos indica que la función de distribución de una variable aleatoria Log-Normal en el punto “ t ” es igual a la función de distribución de una $N(0,1)$ en el punto $\frac{\ln t - \mu}{\sigma}$.

La **función de Supervivencia** será entonces:

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_x^{\infty} \frac{1}{t} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}} dt \quad (48)$$

Si realizamos los cambios que utilizamos para obtener la expresión (46), tendremos:

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{(\ln at)^2}{2\sigma^2}} \frac{dt}{t}$$

$$S(t) = 1 - F\left(\frac{\ln at}{\sigma}\right) \quad (49)$$

donde, $F(y)$ es la función de distribución acumulativa de una variable normal estándar.

La **función de riesgo $h(t)$** viene dada como sabemos por el cociente entre la función de densidad y la función de supervivencia, por tanto:

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

La tasa de fallo $h(t) = \frac{f(t)}{S(t)}$ tiene una fase inicial creciente y luego tiende a cero (aproximadamente a partir de la mediana) cuando $t \rightarrow \infty$.

Para encontrar la media, primero calculamos el momento de orden r respecto al origen de la variable aleatoria T , para lo cual tenemos en cuenta que:

$$Y = \ln T \quad \Rightarrow \quad T = e^Y$$

y además sabiendo que la función generadora de momentos para una variable aleatoria Y , $N(\mu, \sigma)$ es:

$$g_Y(t) = E\left[e^{tY}\right] = e^{t\mu + \frac{t^2\sigma^2}{2}} \quad (50)$$

Luego el momento de orden r es:

$$E[T^r] = E[e^{rY}]$$

que coincide con la función generadora de momentos, (50) de la variable aleatoria Y particularizada para $t = r$:

$$E[T^r] = E[e^{rY}] = e^{r\mu + \frac{r^2\sigma^2}{2}}$$

haciendo $r = 1$ tenemos que la media es:

$$E T = e^{\mu + \frac{1}{2}\sigma^2} \quad (51)$$

Para encontrar la varianza, calculamos el momento de orden 2 que se obtiene haciendo $r = 2$,

Luego, la varianza es:

$$\begin{aligned} \text{Var}(t) &= E[T^2] - E T^2 \\ &= e^{2\mu + \frac{1}{2}4\sigma^2} - \left(e^{\mu + \frac{1}{2}\sigma^2} \right)^2 \\ &= e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} \\ \text{Var}(t) &= e^{2\mu + \sigma^2} \cdot e^{\sigma^2} - 1 \end{aligned} \quad (52)$$

1.6 ESTIMACIÓN PUNTUAL Y POR INTERVALOS PARA LOS PARÁMETROS DE FUNCIONES DE SUPERVIVENCIA.

ESTIMACIÓN DE MAXIMO VEROSIMIL.

Teóricamente cualquier distribución para la cual $S(0) = 1$, puede utilizarse como distribución de supervivencia. Sin embargo, existen ciertas familias de distribuciones específicamente útiles para ajustarse a los datos de un problema de análisis de supervivencia.

Procedemos a la estimación de los parámetros característicos de cada distribución, por máxima verosimilitud (donde se toma como estimación de un parámetro aquel valor que hace lo más probable la muestra obtenida). Para comprobar si el ajuste de la distribución, con esos parámetros, a los datos reales es o no satisfactoria puede recurrirse a paquetes estadísticos que suelen comprobar la *bondad del ajuste* mediante prueba χ^2 .

Sea T una variable aleatoria. La función de densidad de T puede contener uno o varios parámetros. Si la variable aleatoria es continua, con función de densidad que depende de un solo parámetro θ , $f(t;\theta)$, entonces la *función de verosimilitud*, L , de una muestra de “n” valores t_1, t_2, \dots, t_n representa la probabilidad de obtener esos valores al realizar “n” veces una toma de datos. Esta función va a depender del parámetro θ ; así pues, debemos hallar un valor θ que maximice el valor de L, este valor estimado de θ , que denotaremos por $\hat{\theta}$, es siempre una solución de la *ecuación de verosimilitud*:

$$\frac{\partial \ln L}{\partial \theta} = 0$$

Si la función de densidad de la variable T depende de k parámetros $\theta_1, \theta_2, \dots, \theta_k$ el método de máxima verosimilitud proporciona una estimación del valor de cada uno de ellos, y dichas estimaciones $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, verifican las ecuaciones de verosimilitud:

$$\frac{\partial \ln L}{\partial \theta_1} = 0, \frac{\partial \ln L}{\partial \theta_2} = 0, \dots, \frac{\partial \ln L}{\partial \theta_k} = 0$$

es decir, existen tantas ecuaciones de verosimilitud como parámetros a estimar.

Para construir la función de verosimilitud a partir de una muestra aleatoria, suponiendo funciones de probabilidad continuas como las que vamos a estudiar a continuación se tiene en cuenta que la contribución a la verosimilitud de un evento de interés en el tiempo t_j es $f(t_j)$, que representa la probabilidad de que el evento de interés ocurra en ese tiempo, mientras que la contribución de una pérdida en el tiempo t_j será $S(t_j)$, que representa la probabilidad de que el suceso de interés ocurra en un tiempo mayor que t_j . Por tanto, la función de verosimilitud para el conjunto de una muestra formada por n observaciones independientes, de las que d son las ocurrencias del evento de interés en los tiempos t_i ($i = 1, \dots, d$) y m son pérdidas en los tiempos t_j ($j = 1, \dots, m$) será el producto de todas las aportaciones a la verosimilitud en cada momento:

$$L = \prod_{i=1}^d f(t_i) \prod_{j=1}^m S(t_j)$$

que teniendo en cuenta que $h(t) = \frac{f(t)}{S(t)}$, se puede escribir de la forma:

$$L = \prod_{i=1}^d h(t_i) S(t_i) \prod_{j=1}^m S(t_j) = \prod_{i=1}^d h(t_i) \prod_{i=1}^n S(t_i)$$

Ejemplo1. Estimación del Parámetro de la Distribución Exponencial.

Decimos que una variable aleatoria T sigue la ley exponencial si su *función de densidad* $f(t)$ tiene la forma:

$$f(t) = \lambda e^{-\lambda t} \text{ si } t \geq 0, \lambda > 0$$

La correspondiente *función de distribución* $F(t)$, viene expresada por:

$$F(t) = 1 - e^{-\lambda t} \quad t \geq 0$$

La función de Supervivencia, que es la probabilidad de sobrevivir a un instante t, será:

$$S(t) = 1 - F(t) = P(T > t) = e^{-\lambda t} \quad t \geq 0$$

La *función de riesgo*, que da la tasa condicional de fallo o que ocurra el evento de interés de un elemento antes del instante t, está dado por:

$$h(t) = \frac{f(t)}{S(t)} = \lambda \quad t \geq 0$$

es decir, el riesgo es constante a lo largo del tiempo. A esta propiedad de la distribución exponencial como hemos mencionado anteriormente se le suele llamar **pérdida de memoria**.

Para encontrar el estimador de λ partiremos de la expresión general obtenida de la función de máxima verosimilitud,

$$L = \prod_{i=1}^d h(t_i) \prod_{i=1}^n S(t_i)$$

en este caso tenemos que:

$$L = \prod_{i=1}^d \lambda \prod_{i=1}^n e^{-\lambda t_i}$$

su logaritmo es:

$$\ln L = \sum_{i=1}^d \ln \lambda + \sum_{i=1}^n -\lambda t_i = d \ln \lambda - \lambda \sum_{i=1}^n t_i$$

y derivando con respecto a λ :

$$\frac{\partial \ln L}{\partial \lambda} = \frac{d}{\lambda} - \sum_{i=1}^n t_i$$

Igualando a cero se obtiene el estimador de λ :

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^n t_i}$$

Luego, el intervalo de confianza para λ es.

$$\frac{\hat{\lambda} \chi_{2d, 1-\alpha/2}^2}{2d} < \lambda < \frac{\hat{\lambda} \chi_{2d, \alpha/2}^2}{2d}$$

donde $\chi_g^2(t)$ es el cuantil t , de una χ^2 con g grados de libertad.

Cuando d es grande ($d \geq 25$, por ejemplo), $\hat{\lambda}$ se distribuye aproximadamente como una normal con media λ y varianza λ^2/d .

El intervalo de confianza para λ será:

$$\hat{\lambda} - \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{d}} < \lambda < \hat{\lambda} + \frac{\hat{\lambda} Z_{\alpha/2}}{\sqrt{d}}$$

donde Z_t es el cuantil t de una Normal estándar.

Ejemplo 2. Estimación de Parámetros de la Distribución Weibull.

Se trata de una generalización de la distribución exponencial. Decimos que una variable aleatoria T sigue la distribución de Weibull si su *función de densidad* $f(t)$ tiene la forma:

$$f(t) = \frac{\alpha}{\theta^\alpha} t^{\alpha-1} e^{-t/\theta}^\alpha$$

Si realizamos el siguiente cambio $\beta = \frac{1}{\theta^\alpha}$, para evitar el fraccionamiento, tendremos la nueva expresión:

$$f(t) = \alpha \beta t^{\alpha-1} e^{-\beta t^\alpha} \quad \text{si } t \geq 0$$

En la expresión anterior al parámetro β se le llama parámetro de escala, mientras que α se conoce como parámetro forma.

Para esta función de densidad, la función de distribución es:

$$F(t) = \int_0^t f(s) ds = 1 - e^{-\beta t^\alpha}, \quad t \geq 0$$

La función de supervivencia, que da la probabilidad de sobrevivir a un instante t , será:

$$S(t) = P(T > t) = 1 - F(t) = e^{-\beta t^\alpha}, t \geq 0$$

la función de riesgo $h(t)$ que da la tasa condicional de fallo u ocurrencia del evento antes del instante t , viene dada por la expresión:

$$h(t) = \frac{f(t)}{S(t)} = \alpha\beta t^{\alpha-1}, t \geq 0$$

Derivando $h(t)$ con respecto a t

$$h'(t) = \frac{d}{dt} h(t) = \alpha\beta \alpha - 1 t^{\alpha-2}, t \geq 0$$

Las expresiones obtenidas para $h(t)$ y $h'(t)$ nos dan información sobre el tipo de función de tasa de riesgo que puede expresarse para cualquier parámetro de forma α .

1. Cuando $\alpha = 0$, $h'(t)$ y $h(t)$ son siempre cero, luego α no puede ser cero para una situación donde el evento de interés es un hecho cierto.
2. Cuando $\alpha = 1$, $h'(t) = 0$ y $h(t) = \beta$. Esto es una tasa constante de riesgo luego la densidad de fallo o que el evento de interés es $f(t) = \beta e^{-\beta}$ que es una densidad exponencial.
3. Cuando $0 < \alpha < 1$, $h'(t) < 0$. Esto supone que la tasa de riesgo decrece a medida que pasa el tiempo.
4. Cuando $\alpha > 1$, $h'(t) > 0$. Esto implica que la tasa de riesgo aumenta a medida que pasa el tiempo.

Recordemos que la distribución exponencial es útil cuando la tasa de riesgo es constante y no depende del tiempo; de forma que, en el seguimiento de un individuo, la probabilidad de que le ocurra el evento de interés, en los sucesivos intervalos de tiempo, no cambia.

Para encontrar el estimador de β , ocuparemos la función de máxima verosimilitud, teniendo que:

$$L = \prod_{i=1}^d h(t_i) \prod_{i=1}^n S(t_i)$$

en este caso tendremos que.

$$L = \prod_{i=1}^d \alpha \beta t_i^{\alpha-1} \prod_{i=1}^n e^{-\beta t_i^\alpha} = \alpha^d \beta^d \prod_{i=1}^d t_i^{\alpha-1} \prod_{i=1}^d e^{-\beta t_i^\alpha}$$

y su logaritmo es:

$$\ln L = d \ln \alpha + d \ln \beta + (\alpha - 1) \sum_{i=1}^d \ln t_i - \beta \sum_{i=1}^n t_i^\alpha$$

derivando e igualando a cero sus derivadas parciales con respecto a α y β , encontramos los estimadores para α y β :

$$\frac{\partial \ln L}{\partial \beta} = \frac{d}{\beta} - \sum_{i=1}^n t_i^\alpha = 0 \text{ si y solo si } \frac{d}{\beta} = \sum_{i=1}^n t_i^\alpha$$

luego

$$\hat{\beta} = \frac{d}{\sum_{i=1}^n t_i^\alpha}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{d}{\alpha} + \sum_{i=1}^d \ln t_i - \beta \sum_{i=1}^n t_i^\alpha \ln t_i$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{d}{\alpha} + \sum_{i=1}^d \ln t_i - \frac{d}{\sum_{i=1}^n t_i^\alpha} \sum_{i=1}^n t_i^\alpha \ln t_i = 0$$

que es una ecuación no lineal en α , y en consecuencia se puede resolver mediante un método iterativo como el método de Newton – Raphson.

Los intervalos de confianza vendrán dados por las expresiones siguientes:

$$\hat{\alpha} - Z_{\alpha/2} \sqrt{\text{var}(\hat{\alpha})} < \alpha < \hat{\alpha} + Z_{\alpha/2} \sqrt{\text{var}(\hat{\alpha})}$$

$$\hat{\beta} - Z_{\alpha/2} \sqrt{\text{var}(\hat{\beta})} < \beta < \hat{\beta} + Z_{\alpha/2} \sqrt{\text{var}(\hat{\beta})}$$

1.7 MODELOS PARA PRUEBAS ACELERADAS.

Los ensayos acelerados surgen debido a que en ocasiones algunos productos o estudios tienen duraciones tan elevadas que es imposible seguir el experimento hasta el final. Por ejemplo: En componentes diseñados para durar 40 años, es muy improbable que ocurra el evento de interés (Fallo) en el tiempo en que razonablemente se puede realizar el ensayo.

Para ello, se pone a prueba el componente bajo condiciones de trabajo mucho más desfavorables de las habituales y se propicia que el evento de interés se produzca antes del tiempo estimado. La realización de ensayos acelerados es compleja y debe ser planificada por los propios investigadores del estudio, ya que hay que tener en cuenta qué factores hay que acelerar y en que medida.

Un posible esquema de trabajo podría ser el siguiente:

- ✓ Se obtienen datos de tiempos de fallo con diversas aceleraciones.
- ✓ Se estima mediante un análisis Weibull la distribución para cada uno de esos niveles.
- ✓ Se calcula la mediana y los percentiles 10% y 90%.
- ✓ Se dibuja en un gráfico la mediana y los percentiles respecto al nivel de carga.

Para la aplicación de un modelo de vida acelerada se supone que hay una función que depende de las condiciones a las que está sometida la unidad “x” tal que:

$$S(t; x) = P(T \geq t/x) = S_0 t \psi_x$$

donde S_0 es una función de confiabilidad base y $\psi_x \geq 0$.

Ejemplo: Sea T el tiempo de ruptura de un aislador eléctrico sujeto a un voltaje constante “ x ”; suponga que $\psi_x = x^\beta$ para algún β y la función de confiabilidad base es $S_o(t) = e^{-t}$ entonces lo que se está diciendo es que la distribución de T es exponencial con Media del Tiempo de Fallo (MTTF),

$$MTTF = \frac{1}{\psi_x} = x^{-\beta}$$

En los modelos de vida acelerada se puede pensar que el tiempo ‘escalado’ $T\psi_x$ tiene función de Supervivencia S_o ; así que si consideramos la ‘Log vida’

$$\ln T = -\ln \psi_x + \ln W$$

donde W es una variable aleatoria con confiabilidad S_o , o lo que es lo mismo, la distribución de $\ln T$ es la misma que la de $\ln W$ sujeta a un cambio de localización de tamaño $\ln \psi_x$.

La función de riesgo satisface la ecuación:

$$h(t;x) = -\frac{\partial}{\partial t} \log S(t;x)$$

$$h(t;x) = -\frac{\partial}{\partial t} \log S_o(t\psi_x) = \psi_x h_o(t\psi_x)$$

Si estamos en presencia de datos de pruebas aceleradas se puede considerar graficar la ‘log vida’ contra los valores de “ x ”, para obtener información sobre la forma de ψ_x .

Ejemplo:

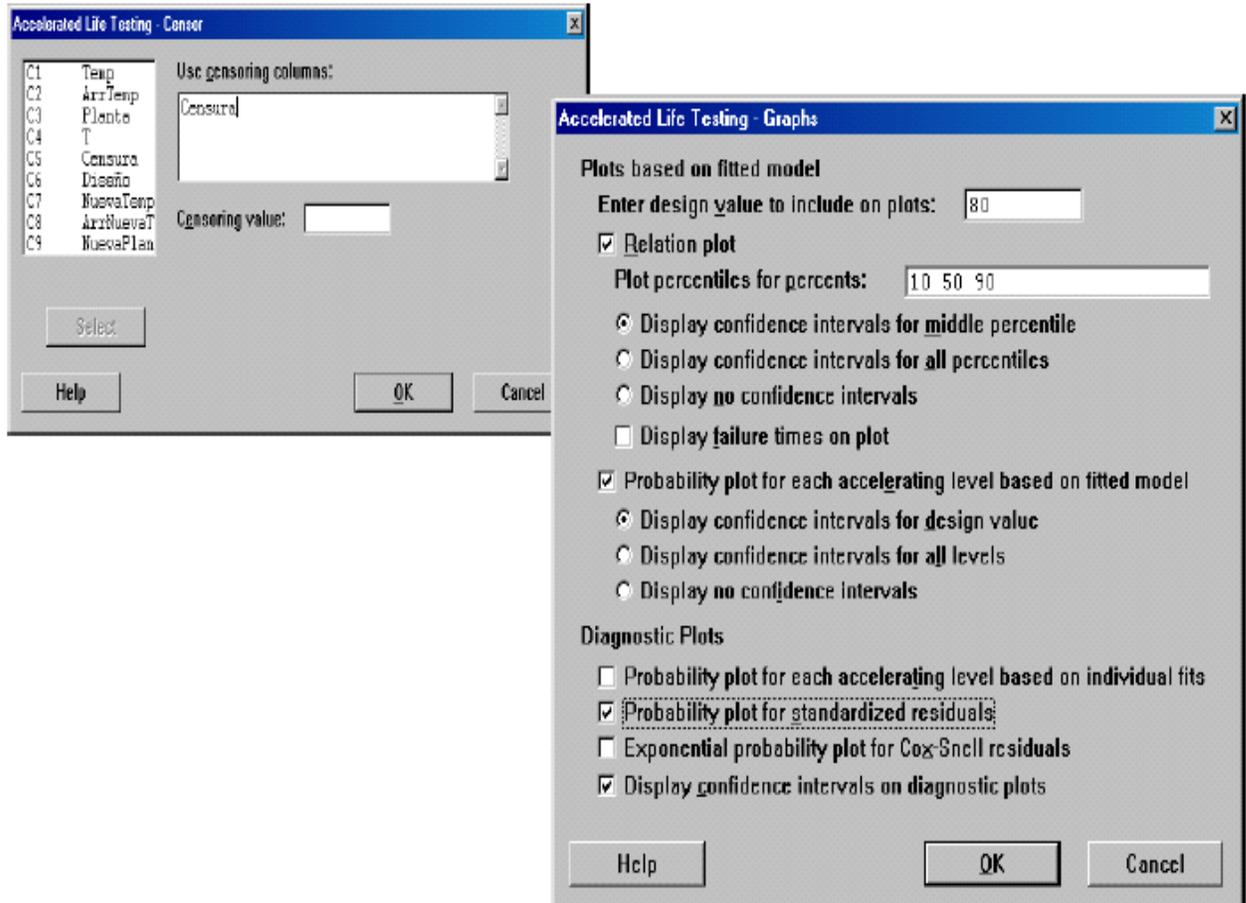
Supóngase que se desea investigar el proceso de deterioro que sufre la capa aislante de los motores eléctricos. Dichos motores soportan habitualmente temperaturas de entre 80° y 100° C Con el fin de minimizar los costos del estudio (tanto de tiempo como de dinero), se llevará a cabo un test de vida acelerada.

En primer lugar, para acelerar el proceso de desgaste, se registrarán los tiempos de deterioro de la capa a temperaturas extremas de 110, 130, 150, y 170° C. Con estos datos, guardados en el archivo **Aislante.mtw**, será posible estimar los tiempos de fallo para las temperaturas habituales de 80 a 100° C:

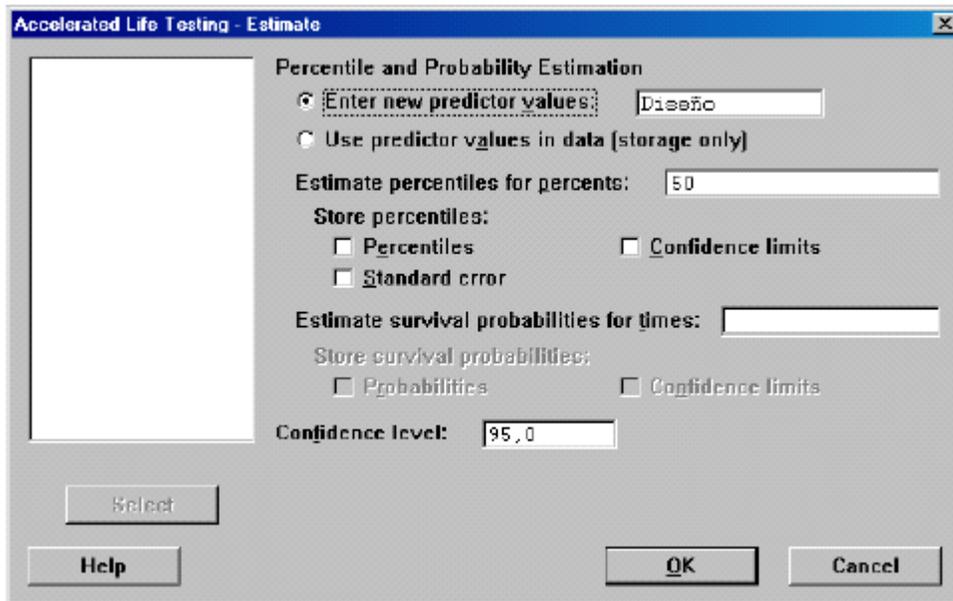
	C1	C2	C3	C4	C5-T	C6	C7	C8	C9
	Temp	ArrTemp	Planta	T	Censura	Diseño	NuevaTemp	ArrNuevaT	NuevaPlanta
1	170	26,1865	1	343	F	80	80	32,8600	1
2	170	26,1865	1	869	F	100	80	32,8600	2
3	170	26,1865	1	244	C		100	31,0988	1
4	170	26,1865	1	716	F		100	31,0988	2
5	170	26,1865	1	531	F				
6	170	26,1865	1	738	F				
7	170	26,1865	1	461	F				
8	170	26,1865	1	221	F				
9	170	26,1865	1	665	F				
10	170	26,1865			C				
11	170	26,1865			C				

Entrada de datos (input): En este caso, se usará una transformación de **Arrhenius** para la variable aceleradora, se tomará $X = Arr(Z)$, siendo Z la variable aceleradora que representa la temperatura (Temp en este ejemplo). En base a un estudio anterior, se sabe también que los tiempos de fallo son ajustables mediante una distribución Weibull.

Será necesario especificar la columna que contiene el indicador de censura (F = fallo, C = censura). Es posible solicitar también, además de los gráficos por defecto, un gráfico de probabilidad para los residuos. Con dicho gráfico se podrá comprobar visualmente si la distribución elegida para los tiempos de fallo (la Weibull en este caso) es correcta:



Finalmente, también se pedirá a **MINITAB** que, a partir del modelo obtenido, realice las predicciones sobre los tiempos de fallo del percentil 50 a las temperaturas de 80 y 100° C (incluidas en la columna Diseño). Es decir, se obtendrán estimaciones del tiempo que tardarían en deteriorarse el 50% de las capas aislantes para una temperatura de 80° C y para una temperatura de 100° C:



La **tabla de regresión** proporciona los coeficientes del modelo. Para una distribución de Weibull (como la que se ha elegido), nuestro modelo sería:

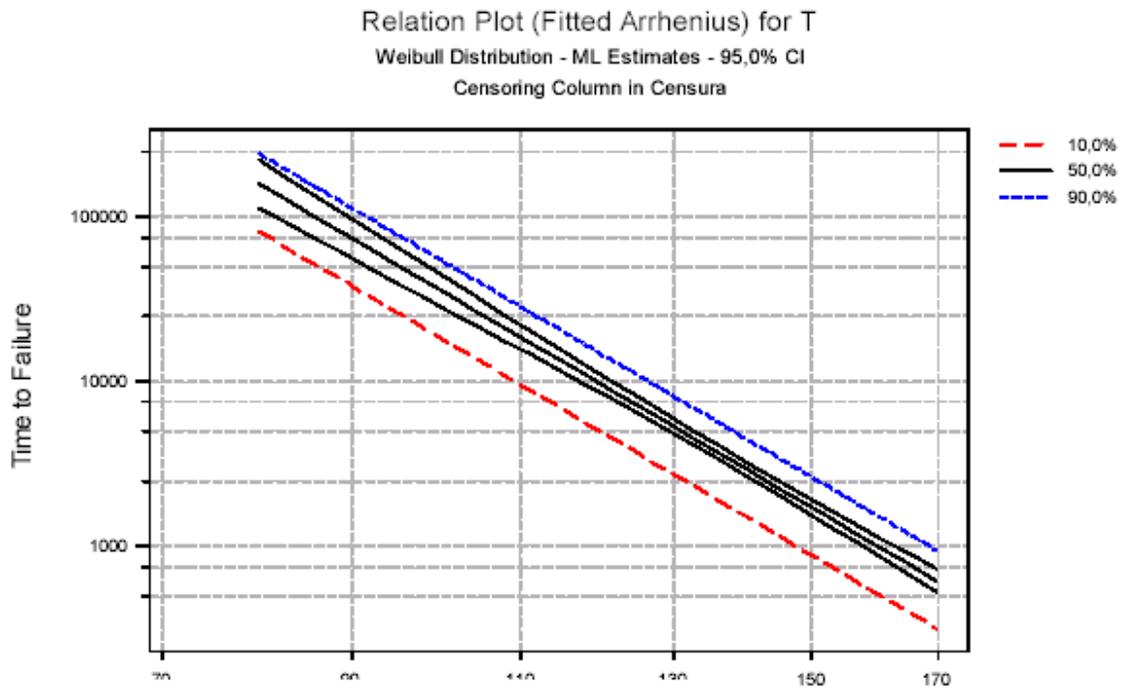
$$\ln T_p = (-15,1874 + 0,83072)Arr(Temp) + \frac{1}{2,8246} \varepsilon_p$$

donde ε sigue una distribución de valores extremos (0,1).

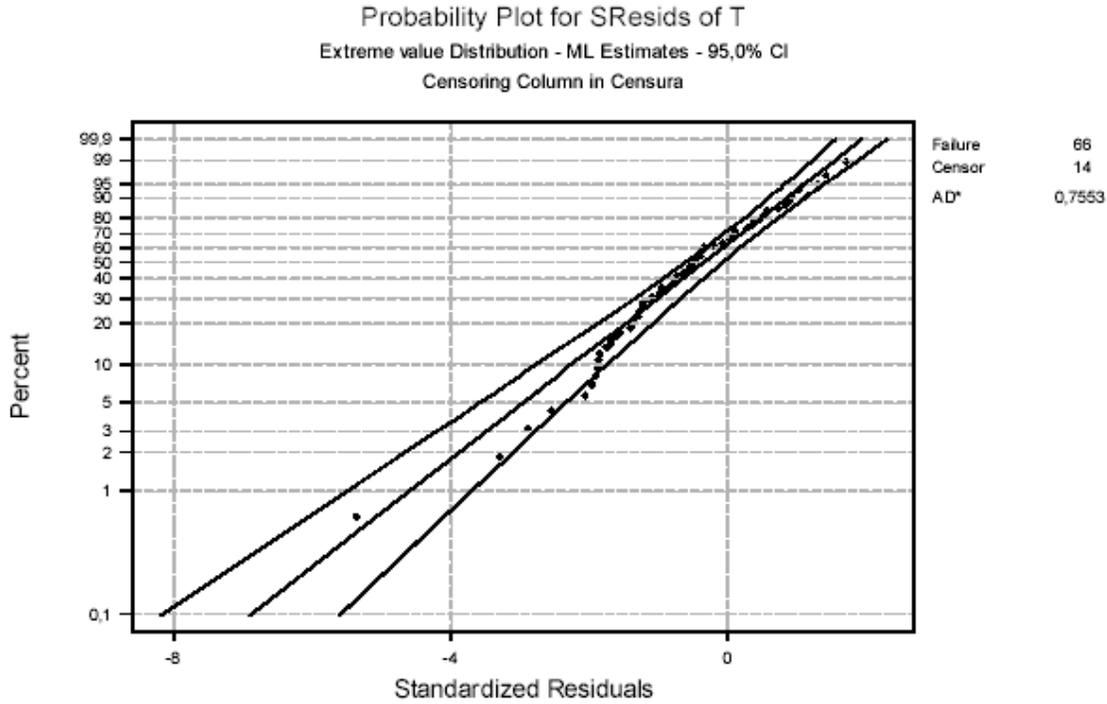
La **tabla de percentiles** muestra los percentiles de orden 50 para las temperaturas solicitadas de 80 y 100° C. El percentil 50 es un buen estimador para el tiempo de duración esperada de una capa de aislamiento sometida a una determinada temperatura. A 80° C, la protección funcionaría alrededor de 159.584,5 horas (o 18,20 años); a 100° C, la misma protección no superaría las 36.948,57 horas (4,21 años).

A partir del gráfico siguiente es posible obtener información sobre la distribución de tiempos de fallo para cada temperatura. En este caso, es posible estimar los percentiles de orden 10, 50, y 90.

Como podemos observar por medio del gráfico de probabilidad de los residuos la distribución elegida se ajusta de forma correcta a las observaciones.



Además como nos muestra la siguiente gráfica los puntos están ubicados muy próximos a la recta central, por lo que podemos concluir que la distribución elegida para modelar el proceso de deterioro que sufre la capa aislante de los motores eléctricos, proporciona una muy buena aproximación bajo las condiciones de temperatura dadas.



CAPITULO II: ESTIMACIÓN NO PARAMÉTRICA DE LA SUPERVIVENCIA.

2.1 INTRODUCCIÓN

El objetivo principal en los problemas de Fiabilidad y Análisis de Supervivencia es la estimación de la función de supervivencia $S(t)$. Esta función es la base para estimar la mayor parte de las funciones y parámetros de interés en el análisis del tiempo de vida.

Si la muestra no contiene observaciones censuradas, la función de supervivencia se estima mediante la función de supervivencia empírica, **fse**, definida como,

$$S(t) = \frac{\text{Nº individuos en la muestra que sobreviven al instante } t}{\text{Nº total de individuos en la muestra}}$$

Este estimador es una función no creciente, toma el valor **1** en todo instante anterior al tiempo de fallo más pequeño y **0** a partir del máximo tiempo de fallo observado; la función permanece constante entre dos instantes de fallo consecutivos y presenta un salto descendente en cada tiempo de fallo observado. Si no hay empates en la muestra, todos los saltos de la función son de altura $1/n$, mientras que si se observan d tiempos de vida iguales a t_i , el salto de $\hat{S}(t)$ en ese instante será de altura d/n .

Cuando en la muestra existen observaciones censuradas la **fse** no es un estimador adecuado porque tiende a subestimar la función de supervivencia. En efecto, el utilizar este estimador es equivalente a considerar que todos los individuos censurados fallan en el instante de censura. Dado que es posible que alguno de los individuos con tiempo de censura menor que t esté vivo en dicho instante, será necesario introducir alguna modificación en el estimador para evitar ese sesgo.

Los métodos estadísticos para estimar los parámetros y funciones de la distribución del tiempo de vida se clasifican en paramétricos y no paramétricos según se basen, en hipótesis específicas sobre la familia a la que pertenece la distribución de T . Las técnicas no paramétricas, que utilizan menos hipótesis, se emplean preferentemente en las primeras fases del estudio cuando se tiene poca información sobre el comportamiento del fenómeno; los resultados obtenidos en estos análisis ayudan a determinar qué distribución de probabilidad representa mejor los datos observados.

En este capítulo se estudian dos procedimientos no paramétricos de estimación a partir de una muestra homogénea: la estimación mediante **Tablas de Vida**, que se basa en datos agrupados en intervalos, y el estimador de **Kaplan-Meier** de la función de supervivencia, que requiere observaciones individuales.

2.2 TABLAS DE VIDA

El tiempo de supervivencia, es una variable continua y por tanto la representación de medidas relacionadas con ella ha de mostrar dicha continuidad, por ejemplo mediante los gráficos a los que nos hemos referido con anterioridad. No obstante, una oportuna discretización de esta variable nos puede permitir elaborar una tabla de gran utilidad práctica. Por discretización de la variable entenderemos aquí una agrupación de los datos en intervalos.

Para crear una tabla de vida procederemos como se explica a continuación. Se subdivide el intervalo temporal de observación desde el punto inicial en intervalos menores, por ejemplo en años. Se contarán las personas que han sobrevivido al menos hasta algún punto de ese intervalo para calcular determinadas probabilidades relacionadas con el momento terminal. Estas probabilidades se utilizarán para calcular o estimar la probabilidad genérica de que una persona viva en un momento determinado.

Por lo dicho anteriormente, observamos dos tipos de dificultades a la hora de efectuar el análisis de datos: el origen de tiempo no es el mismo para los diferentes individuos objeto de estudio y existe ausencia de información, en relación al tiempo de supervivencia, de algunos de los individuos objeto de estudio. Estas dificultades serán salvadas por el conjunto de métodos y técnicas propios del análisis de supervivencia.

Para construir una tabla de vida han de cumplirse las siguientes hipótesis:

- a) Las condiciones experimentales de supervivencia no cambian a lo largo del estudio. Por ejemplo, si se aplica a los pacientes una nueva terapia o medicamento a partir de un momento dado o si se modernizan las condiciones de trabajo de los empleados de la compañía en la cual se trata de estudiar la permanencia de los mismos en la empresa, no podríamos realizar una tabla de vida real.

- b) Una persona/máquina que se comienza a estudiar ahora ha de responder de la misma forma que si se hubiera introducido en el estudio hace, por ejemplo, 5 años.
- c) Las observaciones censuradas no difieren de las que no lo son. De no ser así, esto significaría que las muertes/fallos se producen de manera no aleatoria influenciada por una variable que no se está teniendo en cuenta ahora.

REPRESENTACIÓN SIMBÓLICA DE LOS DATOS.

La matriz de datos del diseño podríamos representarla como la Tabla 2.1, donde una de las variables indicará el tiempo de supervivencia, t , y otra el estado de entidad en el momento del cierre del estudio codificada conforme a criterios del tipo: 1 – muerte/fallo y 0 – censura. Cada uno de los n individuos o unidades estadísticas que son objeto de estudio corresponden a un caso y viene representado en una fila. Los datos del problema consisten en $n(k+2)$ observaciones, n por cada una de las $k+2$ variables o atributos que describen el contexto de la investigación y que caracterizan a los individuos objeto de estudio.

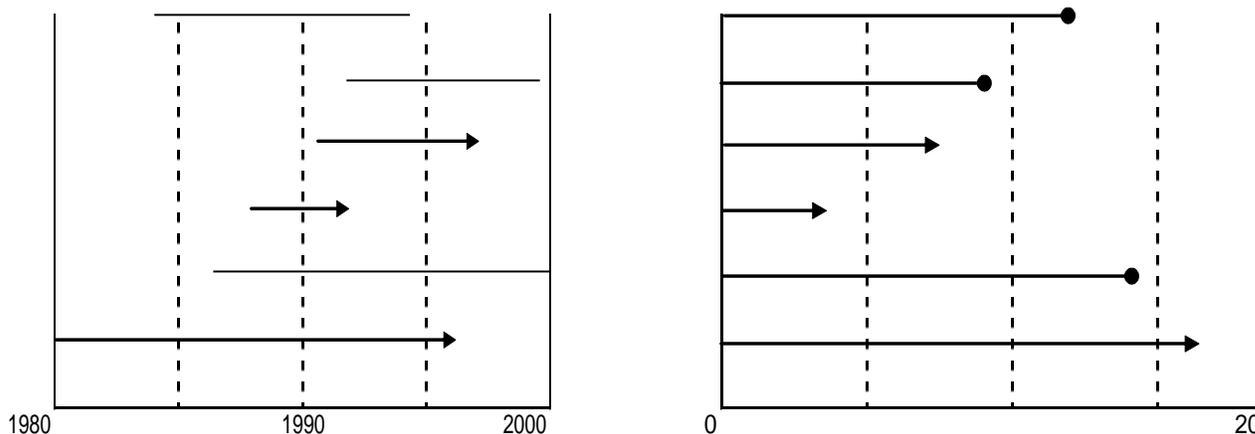
Tabla 2.1: Formato típico de las variables y de los casos

		VARIABLES					
Casos	t	Δ	1	2	3	...	K
1	t_1	δ_1	X_{11}	X_{12}	X_{13}	...	X_{1k}
2	t_2	δ_2	X_{21}	X_{22}	X_{23}	...	X_{2k}
.
n	t_n	δ_n	X_{n1}	X_{n2}	X_{n3}	...	X_{nk}

Una representación típica de los datos observados es la que se muestra en la Tabla 2.1. La representación puede hacerse en tiempo real como se muestra en el gráfico de la izquierda a lo largo de los años 1980 a 2000, con momentos iniciales y terminales reales.

En el gráfico de la derecha la representación se hace poniendo a cero todos los momentos iniciales. En ambos casos una “►” indicará que el suceso (muerte) ha ocurrido, mientras que “●” indicará que el dato está censurado. De esta forma, utilizando el segundo gráfico podemos hacer una división del tiempo en intervalos de cinco años.

Figura 2.1: Representación de tiempos de vida



Los datos correspondientes al gráfico de la derecha se muestran en la Tabla 2.2.

Tabla 2.2: Tiempos de vida correspondientes a la Figura 2.1.

Tiempo	Estado
16	1
13	0
4	1
6	1
9	0
12	0

Si se realiza la tabla de vida con el programa estadístico SPSS, será necesario proporcionar la variable tiempo de supervivencia, así como la variable estado, especificando el valor correspondiente a la ocurrencia del suceso muerte/fallo. La variable tiempo puede estar medida en cualesquier unidades y ha de tener siempre valores positivos. Dichos valores corresponden al tiempo de supervivencia de los no censurados y al tiempo de seguimiento para los censurados.

En el programa SPSS la ocurrencia del suceso (muerte/fallo) viene dada por un valor/valores de la(s) variable(s) estado, de modo que por exclusión todos los demás casos se consideran censurados. Si algún tiempo está ausente (missing), ese caso debería considerarse de modo especial en el estudio. Del mismo modo si alguna censura ha sido producida por motivos relacionados con el estudio, debería tratarse, como ya se ha comentado, con técnicas de análisis de datos perdidos.

Para hacer la tabla de vida es necesario definir la longitud de los intervalos y el momento final de estudio. Puesto que a partir de la tabla pueden construirse las distintas funciones de supervivencia y riesgo, su construcción suele ser una opción gráfica de los programas estadísticos, que en particular el SPSS proporciona como lo podemos ver en la siguiente Tabla:

Tabla 2.2: Tabla de vida

t_{i-1}	n_i'	c_i	n_i	d_i	q_i	p_i	S_i	f_i	H_i	ES	Ef	EH
0	6	0	6.0	1	0.1667	0.8333	0.8333	0.0333	0.0364	0.1521	0.0304	0.0362
5	5	1	4.5	1	0.2222	0.7778	0.6481	0.0370	0.0500	0.2017	0.0334	0.0496
10	3	2	2.0	0	0.0000	1.0000	0.6481	0.0000	0.0000	0.2017	0.0000	0.0000
15	1	0	1.0	1	1.0000	0.0000	0.0000	0.1296	0.4000	0.0000	0.0403	0.0000

La mediana del tiempo de supervivencia es 16.14

Si partimos en cuatro intervalos el gráfico de la derecha de la Figura 2.1 y tomamos el límite inferior obtendremos la primera columna de la Tabla 2.2, además al saber cuantos casos entran en cada intervalo obtendremos la segunda columna y si sabemos cuantos han sido censurados en cada intervalo obtendremos la tercera columna, la cuarta columna son los expuestos a riesgo, mientras que la quinta columna representa los sucesos terminales, el resto de columnas representan cálculos estadísticos cuya explicación la describiremos más adelante.

A partir de ahora supondremos que los datos vienen representados como en el segundo gráfico de la Figura 2.1, donde se presentan los tiempos que median entre el inicio de la observación y el momento en que se produce la muerte o la censura, habría que precisar que los intervalos en cada columna se entienden abiertos por la izquierda y cerrados por la derecha, en otras palabras, en cada intervalo se considera el punto inicial, pero no el punto final.

El límite superior del último intervalo se corresponde aproximadamente con el momento en el que se interrumpe el estudio, aunque para algunos individuos no se haya producido el suceso. Así pues, supondremos que el periodo de observación se divide en k intervalos de tiempo, de t_0 a t_1 el primero, de t_1 a t_2 el segundo,... y de t_{k-1} a t_k el último, así al intervalo correspondiente a la población i lo llamaremos $I_i = t_{i-1}, t_i$, donde está incluido t_{i-1} , pero no t_i , como ya se apuntó anteriormente.

Supondremos que la longitud de cada uno de los intervalos se mantiene fija a lo largo del estudio, aunque podría considerarse de otra forma, el punto medio del intervalo i -ésimo será t_{mi} . Su conocimiento es necesario, pues las funciones de supervivencia y de riesgo serán representadas en estos puntos.

La amplitud o longitud del intervalo se llamará $b_i = t_i - t_{i-1}$, y se utilizará para la estimación de las funciones de densidad y de riesgo, en cada intervalo pueden darse cuatro casos posibles para las observaciones que entran en él. Para ilustrar los casos tomaremos como ejemplo las observaciones en el segundo intervalo del segundo gráfico de la Figura 2.1, comenzando la numeración desde abajo hacia arriba.

- a) Se produce el suceso dentro del intervalo (observación 4): **Observaciones terminales en el intervalo.**
- b) No se produce el suceso en el intervalo, pero sí en algún intervalo posterior (observación 1): **Observaciones terminales en un momento posterior al intervalo.**
- c) No se produce el suceso en el intervalo ni en ningún intervalo posterior del estudio (observación 2 y 6): **Observaciones censuradas en un momento posterior al intervalo.**
- d) Se produce la censura de la observación dentro del intervalo (observación 5): **Observaciones censuradas en el intervalo.**

Recordemos que las censuras en los dos últimos casos podrían producirse por ser supervivientes al final del estudio o bien por tratarse de abandonos. Pasamos ahora a detallar lo que representa cada una de las columnas de la tabla de vida dada anteriormente:

- a) **Punto inicial del intervalo (t_{i-1}):** Extremo izquierdo de cada intervalo.
- b) **Entradas en el intervalo n'_i :** Supervivientes hasta el inicio de ese intervalo. Son los que llegaron al intervalo anterior, menos los que se perdieron o murieron en ese intervalo. Así al inicio del tercer intervalo tenemos 3 supervivientes.

- c) **Censurados en el intervalo** c_i : Número de observaciones censuradas en el intervalo. Podríamos distinguir dos casos posibles, aquellos que al final del estudio no han muerto w_i y aquellos que en un momento dado del estudio se han perdido l_i . A estos se les suele llamar abandonos, si el abandono es debido a causas ajenas a la enfermedad, dichos datos se tratarán como censurados, de no ser así, como ya se ha comentado más arriba, deberían tratarse en cada análisis determinado como datos perdidos. Nosotros supondremos siempre que los abandonos no están motivados por el estudio y por lo tanto los trataremos como censurados, de modo que $c_i = l_i + w_i$. Vemos así que en el transcurso del tercer intervalo aparecen 2 casos censurados.
- d) **Expuestos a riesgo** n_i : Número de casos que entran en el intervalo menos la mitad de los censurados en ese intervalo $\left(n_i = n'_i - \frac{c_i}{2} \right)$. Es una estimación del número de los que tienen riesgo de morir en algún momento del periodo observado. Para hacer esta estimación a los censurados se les ha asignado una probabilidad de $\frac{1}{2}$ de morir si se les hubiera observado durante todo el periodo de observación, se está suponiendo que el tiempo de permanencia en el intervalo se distribuye uniformemente. En el segundo intervalo dicha estimación será por tanto $n_2 = 5 - 1/2 = 4.5$.
- e) **Sucesos terminales** d_i : Número de sucesos en ese intervalo, es decir, de observaciones terminales en el intervalo, por tanto $n'_i = n'_{i-1} - c_{i-1} - d_{i-1}$. Así, mientras que en el tercer intervalo no se produce ninguna baja (muerte/fallo), en los otros tres se da una baja en cada uno.

- f) **Proporción de sucesos terminales** $\left(q_i = \frac{d_i}{n_i} \right)$: Es una estimación de la probabilidad de que un paciente que entra en el intervalo muera dentro de él. Puesto que en el tercer intervalo no hay ninguna baja, dicha probabilidad se hace cero, esto no es lo habitual cuando se dispone de una mayor cantidad de datos.
- g) **Proporción de supervivientes** $p_i = 1 - q_i$: Estimación de la probabilidad de supervivencia de un paciente que entra en un intervalo.
- h) **Proporción acumulada de supervivientes** s_i : Estimación de la probabilidad de supervivencia al final del intervalo i -ésimo. Es un estimador de la función de supervivencia en el instante t_i y nos referimos a menudo a ella como la razón de supervivencia acumulada y se define como $S_0 = 1$ y $S_i = p_i S_{i-1}$, $i = 1, 2, \dots, k$. La supervivencia acumulada en el segundo intervalo será $0,8333 * 0,7778 = 0,6481$, que corresponde a la probabilidad de que sobreviva hasta el final del primer intervalo y hasta el final del último. Puesto que hasta el final del cuarto intervalo no llega ningún paciente, la probabilidad de supervivencia al final de este intervalo es nula, es decir, la probabilidad de sobrevivir más de 20 meses es cero. Con estas estimaciones representaremos después la función de supervivencia.

- i) **Densidad de probabilidad** f_i : Estimación de la probabilidad del suceso (muerte/fallo) por unidad de tiempo, es decir, de la función de densidad del tiempo de ocurrencia del suceso, esta estimación se hace de la manera siguiente: $f_i = S_{i-1} - S_i / b_i$ y en nuestro ejemplo correspondería a la probabilidad de morir en un periodo de un año en el intervalo correspondiente. Quiere expresarse así una cierta probabilidad instantánea propia de la densidad de probabilidad, y que estimamos mediante una distribución discreta. En nuestro ejemplo solamente cuatro valores, correspondientes a los cuatro intervalos nos dan una estimación de dicha densidad de probabilidad. Así, si nos centramos en el intervalo de tiempo de 5 a 10 años la probabilidad de que una persona muera en un año cualquiera de ese intervalo será de un 3.7%.
- j) **Razón de riesgo** h_i : Estimación de la probabilidad del suceso (morir) por unidad de tiempo supuesto que el paciente ha sobrevivido hasta el inicio del intervalo. Es una estimación de la función de riesgo en el instante t_i . Esta función ofrece en cada momento una idea del riesgo que un paciente tiene de morir “en breve”. Esta función era el cociente entre la densidad y la supervivencia. Un estimador del riesgo en cada intervalo viene dado por la relación: $h_i = f_i / 0,5 S_{i-1} + S_i$. Teniendo en cuenta que S_i era la supervivencia al final del intervalo, para calcular el riesgo hemos tomado la supervivencia media del intervalo.

- k) **Error típico de la Supervivencia Acumulada (ES):** Estimación del error típico de la estimación S_i , que vendrá expresado de la siguiente manera:

$$ES \ S_i \approx S_i \sqrt{\sum \left(\frac{d_i}{n_i (n_i - d_i)} \right)}.$$

- l) **Error típico de la densidad de probabilidad (EF):** Estimación del error típico de la estimación f_i .
- m) **Error típico de la razón de riesgo (EH):** Estimación del error típico de la estimación H_i .

Al final de la tabla se suele proporcionar también la mediana del tiempo de supervivencia, que en el ejemplo es 16.14. Esto significa que el 50% de los pacientes considerados en la muestra sobrevive 16.14 años o más. Mientras que casi un 85% de los pacientes sobrevive los primeros 5 años, aproximadamente un 65% llega a los 15 años. Por estar algunos datos censurados esta no es la simple mediana de los números que aparecen en la variable tiempo, que en el ejemplo sería 10.5. Se han de tener en cuenta los datos que están censurados. De no hacerlo así estaríamos actuando como si los datos censurados correspondieran a muertes y por tanto se produciría un sesgo hacia la izquierda. Un estimador de la mediana correcta será el punto temporal para el cual la función de supervivencia vale 0.5.

Como es habitual en el cálculo de la mediana, se utiliza interpolación lineal en el intervalo en el que ha de estar contenida. Podría ocurrir que la función de supervivencia al final del último intervalo no llegara a 0.5. Esto significa que la mediana no puede calcularse, es decir, hay demasiados datos censurados. En estos casos los programas informáticos suelen identificar el comienzo del último intervalo para señalar que la mediana excede dicho valor. El programa SPSS proporciona en este caso el punto inicial del último intervalo con un signo + para indicar que la mediana está más adelante.

2.3 TABLAS DE VIDA PARA VARIOS GRUPOS

Si los datos están agrupados en varias categorías podría interesar un análisis comparativo de las funciones de supervivencia. Para ello se puede construir una tabla de vida adecuada. Supongamos que tenemos los tiempos de permanencia de 43 empleados en una empresa textil, agrupados por sexo. Estamos, por tanto, considerando la influencia del factor sexo.

Disponemos además de la edad en la que entraron a formar parte de la empresa. Aunque de momento no haremos uso de esta variable, es importante comentar que el momento en que se mide la edad es de gran importancia. No sería razonable computar la edad actual de los que siguen en la empresa y la edad en que la dejaron los que ya no forman parte de la misma. Sin embargo, es muy factible encontrar una base de datos de este estilo. En un caso así podríamos computar fácilmente la edad en que fueron contratados, sin más que restar a todos el tiempo de permanencia en la empresa de la edad que figura en la base de datos. Bien es cierto que dicha variable podría haberse tratado directamente como una variable que depende del tiempo y de la posible censura.

En el Ejemplo dispondremos también de la variable coeficiente de inteligencia que, en principio, no dependería del tiempo. Analicemos los tiempos de supervivencia de ambos sexos sin tener en cuenta, de momento, las otras dos variables.

Se producirán dos tablas de vida, del mismo modo como se hizo antes, una para cada grupo. Esto nos permitirá dibujar las funciones de supervivencia, densidad y riesgo en cada caso y llevar a cabo de este modo comparaciones gráficas. Las tablas que hemos obtenido se muestran a continuación:

Tabla 2.3: Tabla de vida para los hombres.

SEXO = 1 (hombres)

t_{i-1}	n'_i	c_i	n_i	d_i	q_i	p_i	S_i	f_i	H_i	ES	EF	EH
0	23.0	4	21.0	4	0.1905	0.8095	0.8095	0.0190	0.0211	0.0857	0.0086	0.0105
10	15.0	5	12.5	4	0.3200	0.6800	0.5505	0.0259	0.0381	0.1217	0.0110	0.0187
20	6.0	2	5.0	0	0.0000	1.0000	0.5505	0.0000	0.0000	0.1217	0.0000	0.0000
30	4.0	2	3.0	0	0.0000	1.0000	0.5505	0.0000	0.0000	0.1217	0.0000	0.0000
40	2.0	2	1.0	0	0.0000	1.0000	0.5505	0.0000	0.0000	0.1217	0.0000	0.0000

La mediana del tiempo de supervivencia es 40.00+

Tabla 2.4: Tabla de vida para las mujeres.

SEXO = 2 (mujeres)

t_{i-1}	n'_i	c_i	n_i	d_i	q_i	p_i	S_i	f_i	H_i	ES	EF	EH
0	20.0	3	18.5	6	0.3243	0.6757	0.6757	0.0324	0.0387	0.1088	0.0109	0.0155
10	11.0	4	9.0	4	0.4444	0.5556	0.3754	0.0300	0.0571	0.1272	0.0122	0.0274
20	3.0	1	2.5	0	0.0000	1.0000	0.3754	0.0000	0.0000	0.1272	0.0000	0.0000
30	2.0	1	1.5	0	0.0000	1.0000	0.3754	0.0000	0.0000	0.1272	0.0000	0.0000
40	1.0	1	0.5	0	0.0000	1.0000	0.3754	0.0000	0.0000	0.1272	0.0000	0.0000

La mediana del tiempo de supervivencia es 15.85.

En las tablas vemos que la mediana para los hombres es superior a 40, mientras que en las mujeres la mediana se sitúa en 15.85. Esto significa que más del 50% de los hombres permanece en la empresa más de 40 años y aproximadamente el 50% de las mujeres no llegan a 16 años. Tanto para los hombres como para las mujeres apenas hay bajas a partir de los 20 años, lo que podría interpretarse como que una vez asentados en la empresa es mucho menor la probabilidad de causar baja. Todo esto, que podría obedecer a una situación razonable, hace pensar que los datos de que disponemos son insuficientes para tomar decisiones significativas.

Interesa ahora comparar la experiencia de supervivencia en la empresa de los hombres y de las mujeres. El contraste de Wilcoxon (Gehan) para el análisis de supervivencia compara ambas experiencias. Para ello introduciremos en el programa estadístico SPSS los valores de ambas tablas de vida, los resultados que se generan son los siguientes:

Comparación de la experiencia de supervivencia usando el estadístico de Wilcoxon (Gehan)

Variable TIEMPO
 Agrupados por SEXO

Estadístico de comparación conjunta = 1.415 G.L. = 1 Prob = 0.2342

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	23	8	15	65.22	3.13604
2	20	10	10	50.00	-3.6000

donde el grupo 1 representa los hombres y el grupo 2 a las mujeres, el total de sucesos terminales esta representado en la tercera columna, el total de censuras en cada grupo se representa en la cuarta columna, mientras que la quinta columna representa el porcentaje de censura por cada grupo, siendo la sexta columna el valor puntual media que como explicaremos más adelante nos servirá como indicador para las diferencias entre las experiencias en estudio.

Como básicamente la puntuación de cada caso no es más que la diferencia entre el número de tiempos mayores que él y el número de tiempos menores, teniendo en cuenta que en algunas ocasiones no hay información suficiente para comparar ambos tiempos y por tanto no tendrá influencia en la puntuación.

Por lo tanto en este caso a pesar de que las puntuaciones en ambos grupos tienen distinto signo, no se pueden asegurar diferencias significativas entre ambas funciones de supervivencia.

Además como podemos ver en los resultados del estadístico de Wilcoxon (Gehan) el contraste es alta, $p = 0.2342$; por tanto, como se mencionó anteriormente para este caso no podemos aceptar diferencias significativas entre ambas experiencias, es decir, entre sus respectivas funciones de supervivencia.

El contraste anterior se basa en las puntuaciones medias para cada grupo, dichas puntuaciones se calculan comparando cada caso con todos los demás del estudio, de modo que la puntuación para un caso crece en un punto si tiene un tiempo de supervivencia mayor que otro y se resta un punto en caso contrario. El estadístico de Wilcoxon (Gehan) se calcula a partir de las puntuaciones de cada uno de los grupos, que finalmente se suele aproximar por una distribución χ^2 .

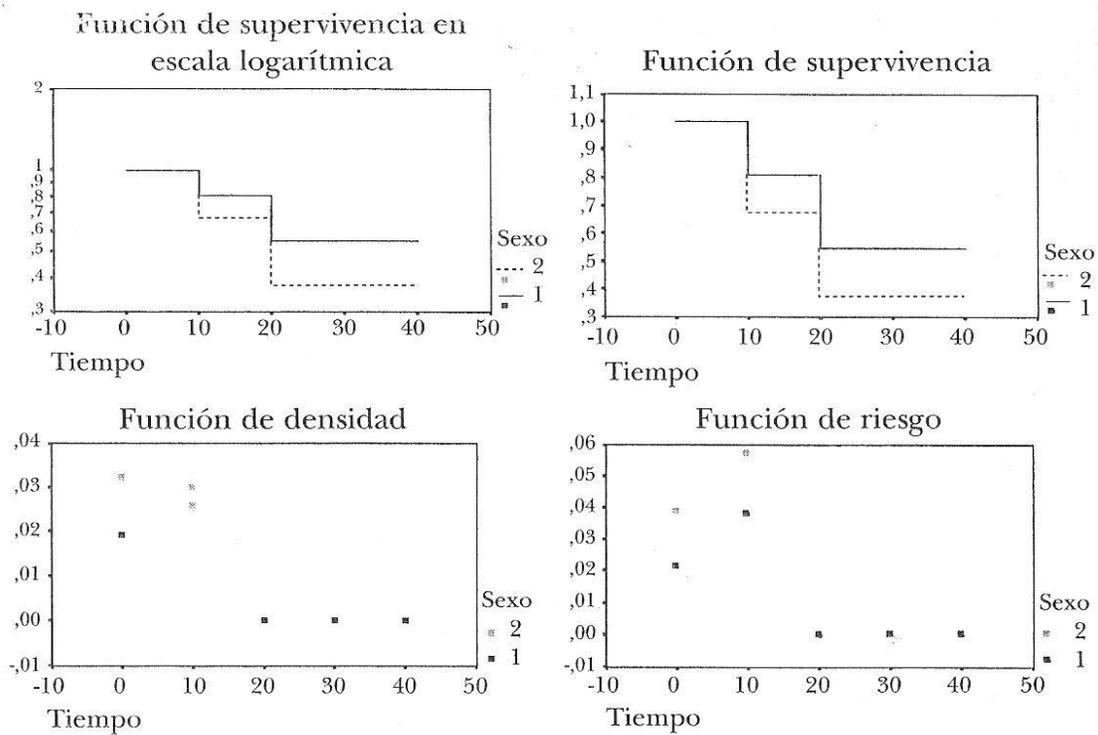
En caso de falta de información suficiente para garantizar que el tiempo de vida sea mayor o menor, debido a los casos censurados, la puntuación no se modifica. Así, por ejemplo, si se compara un paciente que ya ha fallecido con otro que también ha fallecido, bastará contrastar los tiempos correspondientes. Si el mencionado paciente se compara con otro que aún no ha fallecido, no podemos asegurar nada al respecto salvo que este último haya sobrevivido más que el primero hasta el momento. Si el paciente objeto de comparación aún no ha fallecido, entonces, salvo que el otro haya fallecido antes de su tiempo correspondiente, la puntuación no se modificará. Resumimos esto en la tabla siguiente:

Tabla 2.5. Puntaciones para la comparación de funciones de supervivencia.

Caso de referencia:	Otro caso:	Comparación	Modificación
Tiempo t	Tiempos t'	tiempos	Puntuación
Muerto	Muerto	$t < t'$	-1
Muerto	Muerto	$t > t'$	1
Muerto	No Muerto	$t < t'$	-1
Muerto	No Muerto	$t > t'$	0
No Muerto	Muerto	$t < t'$	0
No Muerto	Muerto	$t > t'$	1
No Muerto	No Muerto	$t < t' \text{ ó } t > t'$	0

Las representaciones gráficas vienen ahora superpuestas para hacer las comparaciones de modo visual.

Figura 2.2 Gráficos comparados de las funciones de supervivencia, densidad y riesgo.



En la figura 2.2 pueden observarse las gráficas típicas, aparentemente las gráficas de distintas funciones difieren. En particular, la supervivencia de los hombres en la empresa es mayor que en las mujeres, sin embargo, al hacer el contraste analítico de los datos no hay suficiente evidencia para inferir diferencias significativas. Una vez más los gráficos, aunque ofrecen una primera visualización de la situación, por sí solos no deben tomarse como conclusiones definitivas. Hay que comentar de nuevo que esto se debe en buena medida a la escasez de datos del ejemplo en cuestión.

El programa SPSS ofrece además la posibilidad de introducir un segundo factor, este nuevo factor solamente hace grupos relativos a sus correspondientes niveles y el estudio comentado anteriormente se repite para cada uno de dichos grupos sin hacer comparaciones entre ellos. Si deseáramos realizar comparaciones para las distintas categorías formadas mediante la combinación de ambos criterios tendríamos que construir previamente una nueva variable con dichas categorías.

En el ejemplo considerado el factor solamente tenía dos niveles, para más de dos niveles la situación es completamente análoga. Si el contraste que se lleva a cabo en conjunto la hipótesis nula es que las funciones de supervivencia de cada grupo son iguales, si la hipótesis nula es rechazada será necesario descubrir en que grupos son diferentes y en cuáles no. Los estadísticos de estos contrastes suelen aproximarse por la distribución χ^2 , cuando se dispone de observaciones suficientes, en caso contrario habría que recurrir a contrastes exactos.

Ejemplo:

Los resultados que mostraremos a continuación corresponden a datos de pacientes enfermos con cáncer de pulmón de un Hospital Nacional que han sido sometidos a intervención quirúrgica y cuyo estado de supervivencia ha sido seguido durante cinco años (60 meses). Se dispone de observaciones de 191 pacientes, hemos seleccionado las siguientes variables para nuestro ejemplo:

MUERTE: Muerte del paciente, a lo largo del estudio solamente 67 de los pacientes han fallecido, esta baja mortalidad probablemente está relacionada con el efecto curativo de la cirugía en los pacientes con tumores poco extendidos.

TIEMPO: Tiempo de supervivencia en los no censurados y de seguimiento en los censurados.

FEV1PORC: Función Pulmonar en porcentaje.

EPOC: Categorización de la función pulmonar en baja (1) y alta (2).

ESTADIO: Estadio del cáncer, es decir, lo avanzado y extendido que está, los valores están categorizados de 0 a 6 indicando progresivamente la extensión y gravedad del tumor.

EXTEN: Categorización del estadio del cáncer en bajo (1) y alto (2).

EDAD: Edad del paciente en el momento de la intervención.

AGED: Categorización de la edad en menores de 60 años (1), entre 60 y 70 años (2) y mayores de 70 años (3).

Construimos en primer lugar la tabla de supervivencia para todos los casos sin utilizar ninguno de los factores anteriormente citados. Se han tomado intervalos semestrales para construir la tabla y por tanto la función de supervivencia, los resultados se muestran en la Tabla 2.6 y la función de supervivencia en la Figura 2.3.

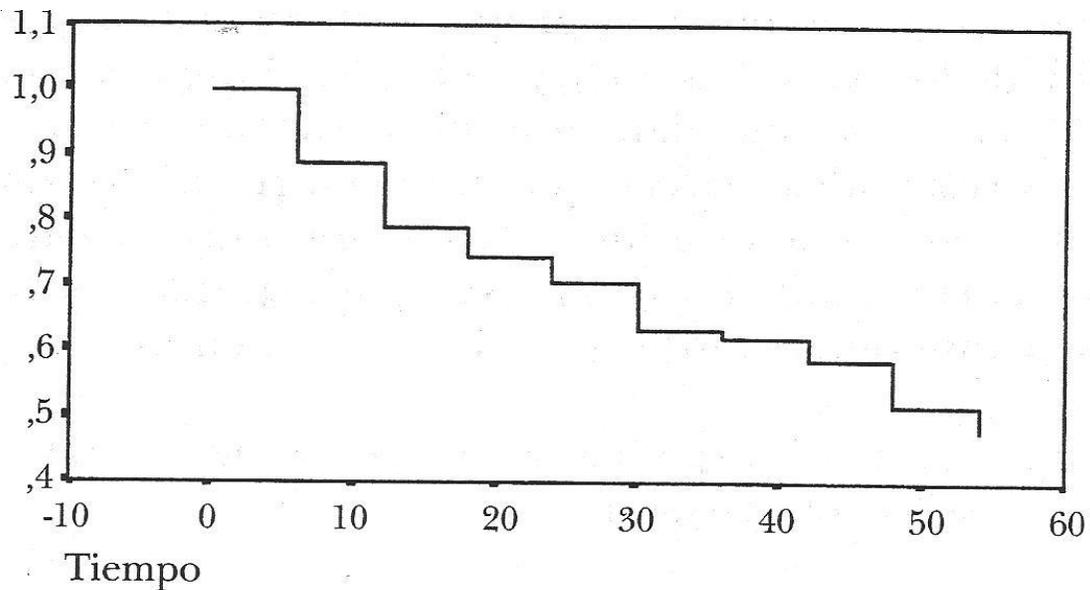
Tabla 2.6 Tabla de vida para enfermos con cáncer de pulmón que han sido intervenidos quirúrgicamente.

t_{i-1}	n_i	c_i	n_i	d_i	q_i	p_i	S_i	f_i	H_i	ES	EF	EH
0	190	2	189.0	21	0.1111	0.8889	0.8889	0.0185	0.0196	0.0229	0.0038	0.0043
6	167	4	165.0	18	0.1091	0.8909	0.7919	0.0162	0.0192	0.0297	0.0036	0.0045
12	145	25	132.5	8	0.0604	0.9396	0.7441	0.0080	0.0104	0.0323	0.0027	0.0037
18	112	24	100.0	5	0.0500	0.9500	0.7069	0.0062	0.0085	0.0347	0.0027	0.0038
24	83	15	75.5	8	0.1060	0.8940	0.6320	0.0125	0.0186	0.0399	0.0042	0.0066
30	60	17	51.5	1	0.0194	0.9806	0.6197	0.0020	0.0033	0.0410	0.0020	0.0033
36	42	9	37.5	2	0.0533	0.9467	0.5867	0.0055	0.0091	0.0450	0.0038	0.0065
42	31	10	26.0	3	0.1154	0.8846	0.5190	0.0113	0.0204	0.0542	0.0062	0.0118
48	18	9	13.5	1	0.0741	0.9259	0.4805	0.0064	0.0128	0.0623	0.0062	0.0128
54	8	8	4.0	0	0.0000	1.000	0.4805	0.0000	0.0000	0.0623	0.0000	0.0000

La mediana del tiempo de supervivencia es 50.96.

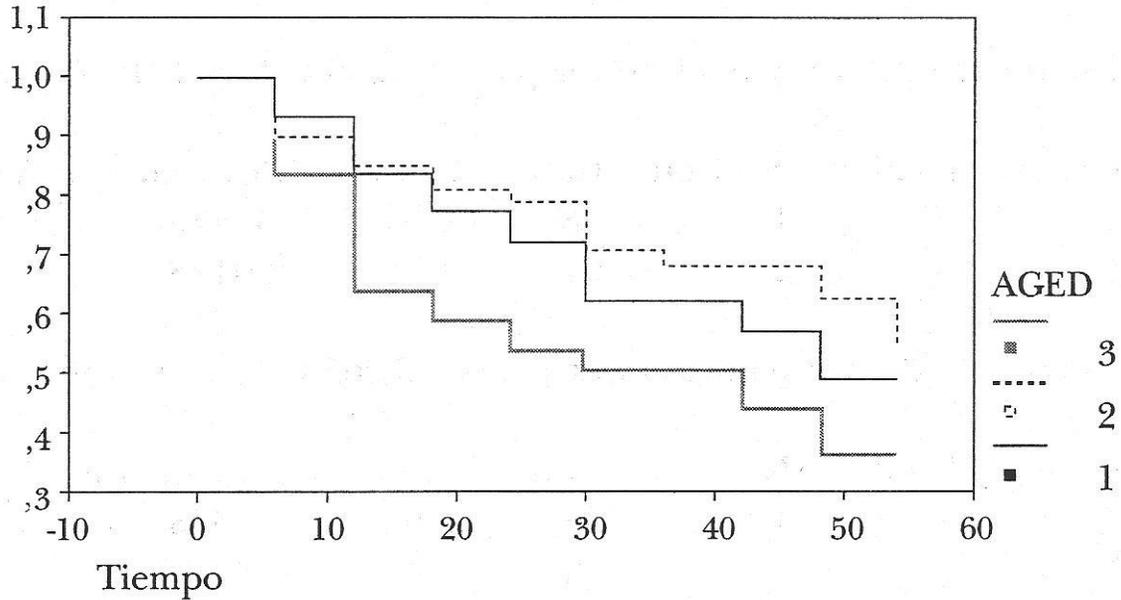
El 50% de los pacientes ha llegado casi a los 60 meses, que es el tiempo que ha durado el estudio, lo cual quiere decir que la supervivencia a lo largo de los 5 primeros años de la intervención quirúrgica es muy grande, siendo más crítica a partir de ese momento, en la Figura 2.3 se observa que la función de supervivencia desciende progresivamente a lo largo de los 5 años hasta caer por debajo del 50%.

Figura 2.3 Función de supervivencia para enfermos con cáncer de pulmón que han sido intervenidos quirúrgicamente.



Analizaremos ahora los tiempos de supervivencia para cada uno de los tres factores mencionados anteriormente, es decir, edad, extensión del tumor y función pulmonar tal y como se han codificado. Mostramos en la Figuras 2.4, 2.5 y 2.6 las funciones de supervivencia correspondientes a cada factor.

Figura 2.4 Funciones de supervivencia para enfermos con cáncer de pulmón según grupos de edad.



Observamos, en la gráfica de la Figura 2.4, diferencias claras para mayores de 70 años, pero no así para los otros dos grupos de edad, donde incluso la función de supervivencia para los más jóvenes es ligeramente inferior. Para el primer grupo de edad, menores de 60 años, tenemos una mediana de supervivencia de 47.27 meses, es decir, casi 4 años, para el segundo grupo de edad más del 50% de los pacientes sobrevive al menos 54.00 meses, mientras que en el tercer grupo de edad, mayores de 70 años, el 50% de los pacientes no llega a los 36.14 meses de supervivencia, es decir, aproximadamente unos 2 años y medio. A la vista de estos resultados, parece que el tercer grupo de edad es un grupo de riesgo, mientras los otros dos son muy similares. Si hacemos los correspondientes contrastes con el programa SPSS para comparar la experiencia de supervivencia usando el estadístico de Wilcoxon (Gehan) entre los tres grupos obtenemos los resultados siguientes:

Comparación de la experiencia de supervivencia usando el estadístico de Wilcoxon (Gehan)

Variable TIEMPO
Agrupados por AGED

Estadístico de comparación conjunta = 7.454 G.L. = 2 Prob = 0.0241

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	57	19	38	66.67	5.7368
2	85	24	61	71.76	12.1882
3	48	24	24	50.00	-28.3958

Estadístico de comparación por pares = 0.250 G.L. = 1 Prob = 0.6172

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	57	19	38	66.67	-3.0000
2	85	24	61	71.76	2.0118

Estos resultados confirman lo que ya apuntábamos antes a la vista del gráfico y de las medianas, la comparación global es significativa entre las funciones de supervivencia. Si contrastamos la función de supervivencia para cada pareja de grupos vemos que para los dos primeros la significación muestral, $P = 0.6172$, es muy alta, lo que nos ofrece una clara evidencia de que la supervivencia en ambos grupos es igual.

Al introducir el tercer grupo al estudio es claramente diverso de los anteriores con significaciones muestrales por debajo del 5%. Como se muestran en los dos resultados siguientes:

Estadístico de comparación por pares = 3.965 G.L. = 1 Prob = 0.0464

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	57	19	38	66.67	8.7368
3	48	24	24	50.00	-10.3750

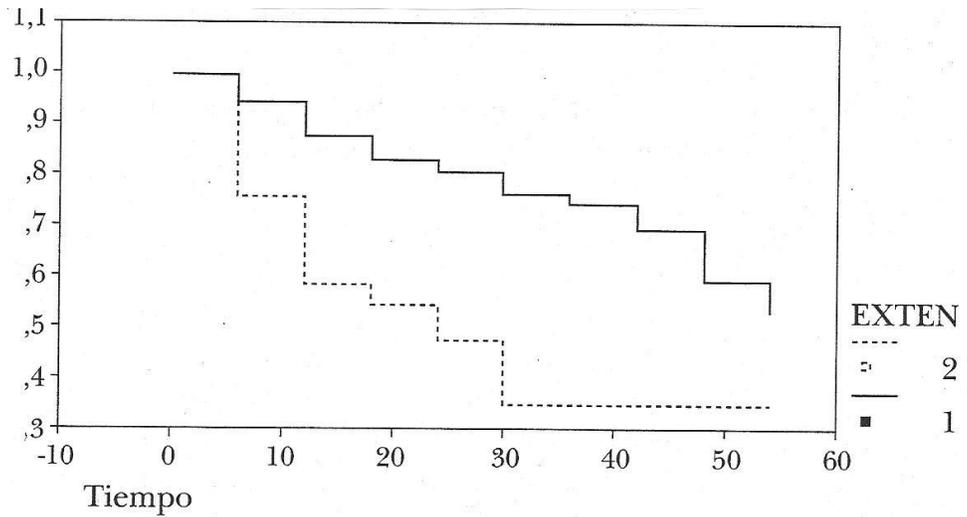
Estadístico de comparación por pares = 6.621 G.L. = 1 Prob = 0.0101

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
2	85	24	61	71.76	10.1765
3	48	24	24	50.00	-18.0208

Por tanto, parece conveniente agrupar las edades en dos grupos, según sean mayores o menores de 70 años, para llevar a cabo el análisis de supervivencia.

Veamos qué ocurre con el factor extensión del tumor, en la Figura 2.5 se observa una clara diferencia entre la supervivencia de enfermos con una extensión mayor o menor del tumor.

Figura 2.5 Funciones de supervivencia para enfermos con cáncer de pulmón según la extensión del tumor.



De igual manera al realizar el contraste de Wilcoxon (Gehan), obtenemos esta misma conclusión con una significación muestral muy baja (Prob = 0.0000), como se ve en el siguiente resultado:

Comparación de la experiencia de supervivencia usando el estadístico de Wilcoxon (Gehan).

Variable Tiempo
 Agrupados por EXTEN

Estadístico de comparación por conjunta = 25.386 G.L. = 1 Prob = 0.0000

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	131	33	98	74.81	20.7252
2	59	34	25	42.37	-46.0169

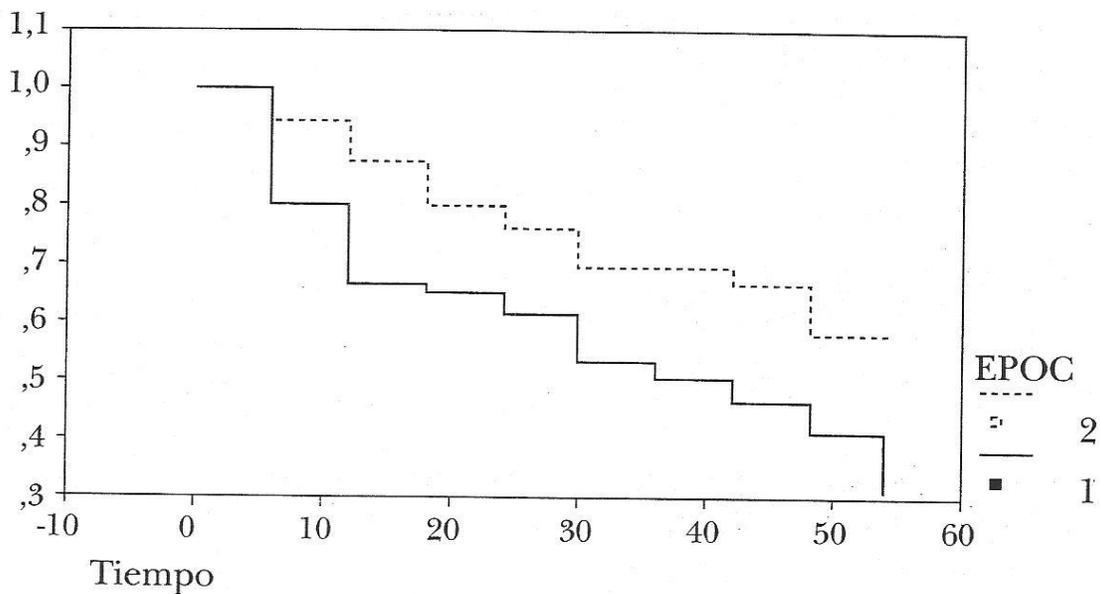
De nuevo la Figura 2.6 nos muestra diferencias significativas entre enfermos con alta y baja función pulmonar. Diferencias, que se ven confirmadas mediante el correspondiente contraste de hipótesis con una significación muestral de $P = 0.0029$.

Variable TIEMPO
Agrupados por EPOC

Estadístico de comparación conjunta = 8.845 G.L. = 1 Prob = 0.0029

Grupo	Total N	Nº Cen	Cen	Pct Cen	Puntuación Media
1	74	36	38	51.35	-22.5000
2	113	31	82	72.57	14.7345

Figura 2.6 Funciones de supervivencia para enfermos con cáncer de pulmón según la función del pulmonar.



2.4 ESTIMACIÓN DE LAS FUNCIONES DE SUPERVIVENCIA

El objetivo de este apartado es estimar e interpretar las funciones de supervivencia, densidad de probabilidad y riesgo, para un conjunto de datos.

Por tanto vamos a:

- a) Describir las funciones de supervivencia.
- b) Utilizar las **tablas de vida** para estimar dichas funciones.
- c) Estimar las funciones de supervivencia, densidad de probabilidad y riesgo, para datos censurados, mediante el método actuarial y el de Kaplan-Meier.
- d) Formular una serie de ejemplos tomados de las ciencias de la vida.

Vamos a considerar el supuesto más fácil posible, en el cual todos los individuos observados mueren durante el período en que está abierto el estudio y no hay censuras, los datos que surgen en estas condiciones son denominados datos de mortalidad completa.

En las ciencias de la vida, la medicina en especial, este tipo de situaciones aparecen en estudios de mortalidad en grupos de personas, cada una de las cuales contrae la enfermedad en torno a la misma edad, la cual se especifica como restricción al comienzo del estudio. Para cada unidad objeto de estudio, la edad en el diagnóstico, suceso inicial, y la edad al fallecer, suceso final, son registradas como datos básicos del estudio. La distancia entre estas dos edades representa el tiempo de supervivencia.

Sean t_1, t_2, \dots, t_n los tiempos de supervivencia de n individuos sujeto a estudio, una buena estimación de la función de supervivencia $S(t)$ es la proporción de individuos de la muestra que sobreviven al instante t ; es decir:

$$S(t) = \frac{N^\circ \text{ individuos en la muestra que sobreviven al instante } t}{N^\circ \text{ total de individuos en la muestra}}$$

Como al comienzo del estudio se mencionó que todos los individuos están vivos y al final todos mueren, se tiene que.

$$S_{t_0} = 1 \quad \text{y} \quad S_{t_n} = 0$$

Si ordenamos de menor a mayor los tiempos de supervivencia de la muestra y les asignamos el rango o puesto que ocupan en dicha ordenación:

$$t_1 < t_2 < \dots < t_n \quad \text{muestra ordenada} \rightarrow t_k \xrightarrow{\text{rango}} k$$

En el caso de no tener observaciones iguales, es decir, todos los tiempos de muerte son diferentes, la función de supervivencia para el tiempo t_i puede estimarse de la forma:

$$S_t = \frac{N^\circ \text{ individuos en la muestra que sobreviven al instante } t}{N^\circ \text{ total de individuos en la muestra}} = \frac{n-i}{n}$$

Recordemos que la función de supervivencia teórica es $S_t = P(T > t)$ y la función de supervivencia empírica para esa muestra de n tiempos $S_n(t)$, es:

$$S_n(t) = \begin{cases} 1 & \text{si } t < t_1 \\ \frac{n-i}{n} & \text{si } t_i < t < t_{i+1} \\ n & \text{si } t \geq t_n \end{cases}$$

Por otra parte la distribución teórica de las muertes, es $F(t) = P(T \leq t) = 1 - S(t)$ y la distribución empírica para esa muestra de n tiempos $F_n(t)$, tiene la siguiente expresión:

$$F_n(t) = \begin{cases} 0 & \text{si } t < t_1 \\ \frac{i}{n} & \text{si } t_i < t < t_{i+1} \\ 1 & \text{si } t \geq t_n \end{cases}$$

En caso de observar empates, es decir, si dos o más tiempos de muertes son iguales, tomaremos el mayor rango de estos t_i . Supongamos que $t_2 = t_3 = t_4$, entonces.

$$S(t_2) = S(t_3) = S(t_4) = \frac{n-4}{n} \text{ y } F(t_2) = F(t_3) = F(t_4) = \frac{4}{n}$$

Ejemplo:

Supongamos que estamos interesados en analizar la supervivencia de los pacientes diagnosticados de cáncer de recto, en un hospital especializado de nuestro país, en dos períodos de tiempo:

- a) años 1935 a 1944
- b) años 1945 a 1954.

Los datos originales, una vez tratados, dieron lugar a las tablas siguientes:

Tabla 2.7: Distribución de frecuencias por intervalos. Ambos períodos.

<i>Intervalo</i>	<i>Individuos</i>	<i>Muertes</i>	<i>Sobrevivientes</i>
1	1131.0	352	779.0
2	767.5	133	634.5
3	623.5	100	523.5
4	513.0	62	451.0
5	439.0	49	390.0
6	356.5	42	314.5
7	269.0	28	241.0
8	210.0	12	198.0
9	173.5	12	161.5
10	136.0	13	123.0

En la Tabla anterior tenemos diez intervalos con un promedio de individuos vivos dentro del periodo de 1935 a 1944, así como la cantidad de muertos en estos intervalos, estos datos nos permitirán obtener los sobrevivientes por intervalo, simplemente restando a los individuos que ingresaron al estudio los fallecidos en cada intervalo; mientras que en la Tabla 2.8 sucede algo similar con la diferencia que se ha agregado la función de supervivencia expresada como la división de los sobrevivientes entre los individuos que ingresan a cada intervalo.

Tabla 2.8: Distribución de frecuencias por intervalos años 1935 a 1944.

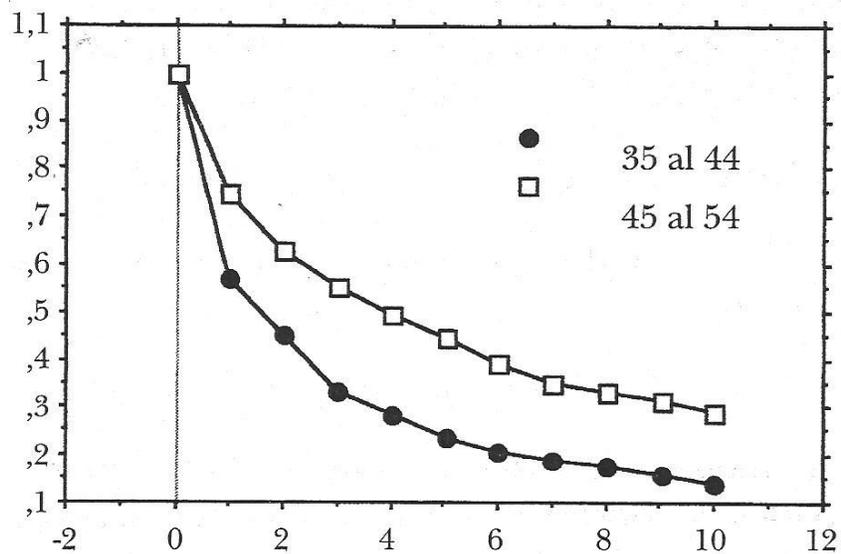
Intervalo	Individuos	Muertes	Sobrevivientes	$S_1 t_i = (4)/(2)$
(1)	(2)	(3)	(4)	
0				1
1	387.0	167	220.0	0.5685
2	218.5	45	173.5	0.4514
3	172.5	45	127.5	0.3336
4	127.0	19	108.0	0.2837
5	108.0	17	91.0	0.2390
6	90.5	11	79.5	0.2100
7	79.0	8	71.0	0.1887
8	71.0	5	66.0	0.1754
9	65.5	6	59.5	0.1593
10	59.0	7	52.0	0.1404

Similar análisis tiene el periodo comprendido entre los años 1945 a 1954, que como se mencionó anteriormente se ha agregado la función de supervivencia e iniciamos nuestro intervalo en “Cero”, que como ya se mencionó anteriormente significa que al comienzo del estudio todos los individuos están vivos y al final todos mueren.

Tabla 2.9: Distribución de frecuencias por intervalos. Años 1945 a 1954.

<i>Intervalo</i>	<i>Individuos</i>	<i>Muertes</i>	<i>Sobrevivientes</i>	$S_2 t_i$
0				1
1	744.0	185	559.0	0.7513
2	549.0	88	461.0	0.6309
3	451.0	55	396.0	0.5539
4	386.0	43	343.0	0.4922
5	331.0	32	299.0	0.4446
6	266.5	31	235.0	0.3928
7	190.0	20	170.0	0.3514
8	139.0	7	132.0	0.3337
9	107.5	6	101.5	0.3151
10	77.0	6	71.0	0.2905

La Figura 2.7 muestra la representación conjunta de las curvas de supervivencia.



Esta representación gráfica parece señalar que la supervivencia en los años 45-54 es mayor que en los años 35-44, aunque debemos recurrir a un contraste más exhaustivo para determinar si esa diferencia que muestra la representación gráfica es o no significativa.

La situación descrita anteriormente, en la cual teníamos datos de mortalidad completa, no es muy común en la vida real. Durante el período de observación de una muestra de individuos, podemos perder unidades objeto de estudio por diversas razones que no sean la propia muerte.

Existen dos métodos principales para la estimación de las curvas de supervivencia, *el actuarial* y *el del producto de Kaplan y Meier*, con el objeto de resolver los problemas planteados por la ausencia de información provocada por las censuras. Ambos métodos estiman la supervivencia para un determinado período de tiempo en función de todas las supervivencias anteriores. La principal diferencia entre ambos radica en que el actuarial define previamente intervalos de tiempo.

2.4.1 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA: MÉTODO ACTUARIAL.

El método actuarial (Bohmer, 1912) es adecuado cuando el número de observaciones es grande. Al principio se definen intervalos de tiempos que van a ser utilizados para estimar la función de supervivencia y es posible que los resultados dependan de la precisión de esta definición de intervalos.

El tiempo de observación va desde $t = 0$ hasta $t = A$, se divide en r intervalos elegidos a priori arbitrariamente:

$$0, a_1, a_1, a_2, \dots, [a_{j-1}, a_j, \dots, a_{r-1}, A$$

Sea P_j la probabilidad de sobrevivir a $[a_{j-1}, a_j]$, sabiendo que se está vivo al comienzo

$$P_j = \Pr T \geq a_j / T \geq a_{j-1}$$

La base de este método consiste en estimar, para cada intervalo, estas probabilidades de sobrevivir P_j . Para ello es necesario determinar el número n_j de unidades expuestas al riesgo de fallecer durante $[a_{j-1}, a_j]$ y el número d_j de unidades que alcanzan el suceso de interés o muerte durante dicho intervalo de tiempo. Se tiene que:

$$P_j = \frac{n_j - d_j}{n_j} = 1 - \frac{d_j}{n_j}$$

Para hallar el número n_j de individuos con riesgo de fallecer durante $[a_{j-1}, a_j]$ calculamos

- 1- Número de individuos vivos al comienzo del intervalo: n'_j
- 2- Número de fallecidos a lo largo del intervalo: d_j
- 3- Número de individuos cuyo seguimiento finaliza dentro del intervalo (censuras): c_j
- 4- Número de individuos vivos al comienzo del siguiente intervalo:
$$n'_{j+1} = n'_j - d_j - c_j.$$

Si en el intervalo $[a_{j-1}, a_j]$ no hay censuras, tenemos que $n_j = n'_j$. Si en el intervalo $[a_{j-1}, a_j]$ hay censuras, tenemos individuos que sólo son observados durante una parte del intervalo.

Estos individuos no están expuestos al riesgo de fallecer durante todo el intervalo, sino tan sólo durante el tiempo que permanecen controlados. (Si un individuo está censurado el día 20, en un intervalo de 30 días, contribuye con $20/30 = 0.67$ al número de individuos expuestos al riesgo y no como 0). El método asume que si c_j es el número de censuras en

$[a_{j-1}, a_j]$, su contribución es $c_j/2$, y:

$$n_j = n'_j - \frac{c_j}{2}$$

se obtiene, en consecuencia, que la probabilidad de sobrevivir a a_j , sabiendo que se ha sobrevivido a a_{j-1} , es:

$$P_j = 1 - \frac{d_j}{n_j - \frac{c_j}{2}}$$

y la **estimación actuarial** de la supervivencia S_{t_i} , es el producto de las probabilidades de sobrevivir al intervalo a_{i-1}, a_i y a todos los intervalos anteriores

$$\hat{S}_{t_i} = P_1 P_2 P_3 \dots P_i = \prod_{j/a_j \leq a_i} \left[1 - \frac{d_j}{n_j - \frac{c_j}{2}} \right]$$

Para todo $t \in [a_{j-1}, a_j]$ se obtiene t por interpolación lineal entre los valores obtenidos en a_{j-1} y en a_j .

$$S_t = S_{a_{j-1}} + \frac{S_{a_j} - S_{a_{j-1}}}{a_j - a_{j-1}} (t - a_{j-1})$$

y para estimar la varianza del estimador actuarial de $S(t)$ se utiliza una fórmula debida a Greenwood (1926).

$$\widehat{Var} \left[\widehat{S}(t) \right] = \widehat{S}(t)^2 \sum_{h/a_h \leq t} \frac{d_h}{n_h(n_h - d_h)}$$

Ejemplo:

En un ensayo (Freireich, 1963) sobre la emisión, en semanas, de un proceso patológico a cuyas unidades objeto de estudio se les sometió a dos tratamientos diferentes:

a) recibir 6-MP y b) recibir un placebo

donde el suceso de interés es la remisión de dicho proceso patológico, se obtuvieron los siguientes tiempos de remisión del proceso:

Grupo 1 (21 individuos), sometido a tratamientos: 6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+.

Grupo 2 (21 individuos), bajo condiciones inocuas: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.

Las tablas 2.10 a) y b) muestran los cálculos de las funciones de supervivencia para ambos grupos mediante el método actuarial.

Tabla 2.10 a): Tabla de cálculo, Grupo 1

$[a_{j-1}, a_j]$	$n'_j = n'_{j-1} - c_{j-1}d_{j-1}$	c_j	d_j	$n_j = n'_j - \frac{c_j}{2}$	$P_j = 1 - \left(\frac{d_j}{n_j}\right)$	$\hat{S}_{a_j} = \hat{S}_{a_{j-1}} P_j$
0,3	21	0	0	21	1.000	1.000
3,6	21	0	0	21	1.000	1.000
6,9	21	1	4	20.5	0.805	0.805
9,12	16	3	1	14.5	0.931	0.749
12,15	12	0	1	12	0.917	0.687
15,18	11	1	1	10.5	0.905	0.622
18,21	9	2	0	8	1.000	0.622
21,24	7	0	2	7	0.714	0.444
24,27	5	1	0	4.5	1.000	0.444
27,30	4	0	0	4	1.000	0.444
30,33	4	2	0	3	1.000	0.444
33,36	2	2	0	1	1.000	0.444

Tabla 2.10 b): Tabla de cálculo, Grupo 2

$[a_{j-1}, a_j]$	$n'_j = n'_{j-1} - c_{j-1}d_{j-1}$	c_j	d_j	$n_j = n'_j - \frac{c_j}{2}$	$P_j = 1 - \left(\frac{d_j}{n_j}\right)$	$\hat{S}_{a_j} = \hat{S}_{a_{j-1}} P_j$
0,3	21	0	5	21	0.762	0.762
3,6	17	0	5	17	0.706	0.538
6,9	12	0	4	12	0.667	0.359
9,12	8	0	2	8	0.750	0.269
12,15	6	0	2	6	0.667	0.179
15,18	4	0	2	4	0.500	0.089
18,21	2	0	2	2	0.000	0.000
21,24	0	0	0	0	1.000	0.000
24,27	0	0	0	0	1.000	0.000
27,30	0	0	0	0	1.000	0.000
30,33	0	0	0	0	1.000	0.000
33,36	0	0	0	0	1.000	0.000

Si queremos saber cuánto vale la función de supervivencia del grupo 1 a tiempo 8 semanas, por interpolación, como $8 \in [6,9]$, tendremos que:

$$S(8) = 0.805 + \frac{0.749 - 0.805}{9 - 6} (8 - 6) = 0.768$$

Si tomamos otros intervalos, por ejemplo de amplitud 5, tendríamos otras funciones de supervivencia, dadas por las Tablas 2.11 a) y b).

Tabla 2.11 a): Tabla de cálculo, Grupo 1

$[a_{j-1}, a_j]$	$n'_j = n'_{j-1} - c_{j-1}d_{j-1}$	c_j	d_j	$n_j = n'_j - \frac{c_j}{2}$	$P_j = 1 - \left(\frac{d_j}{n_j}\right)$	$\hat{S} a_j = \hat{S} a_{j-1} P_j$
0,5	21	0	0	21	1.000	1.000
5,10	21	2	4	20	0.800	0.800
10,15	15	2	2	14	0.857	0.686
15,20	11	2	1	10	0.900	0.617
20,25	8	1	2	7.5	0.733	0.452
25,30	5	1	0	4.5	1.000	0.452
30,35	4	4	0	2	1.000	0.452

Tabla 2.11 b): Tabla de cálculo, Grupo 2

$[a_{j-1}, a_j]$	$n'_j = n'_{j-1} - c_{j-1}d_{j-1}$	c_j	d_j	$n_j = n'_j - \frac{c_j}{2}$	$P_j = 1 - \left(\frac{d_j}{n_j}\right)$	$\hat{S} a_j = \hat{S} a_{j-1} P_j$
0,5	21	0	7	21	0.667	0.667
5,10	14	0	6	14	0.571	0.381
10,15	8	0	4	8	0.500	0.190
15,20	4	0	2	4	0.500	0.095
20,25	2	0	2	2	0.000	0.000
25,30	0	0	0	0	1.000	0.000
30,35	0	0	0	0	1.000	0.000

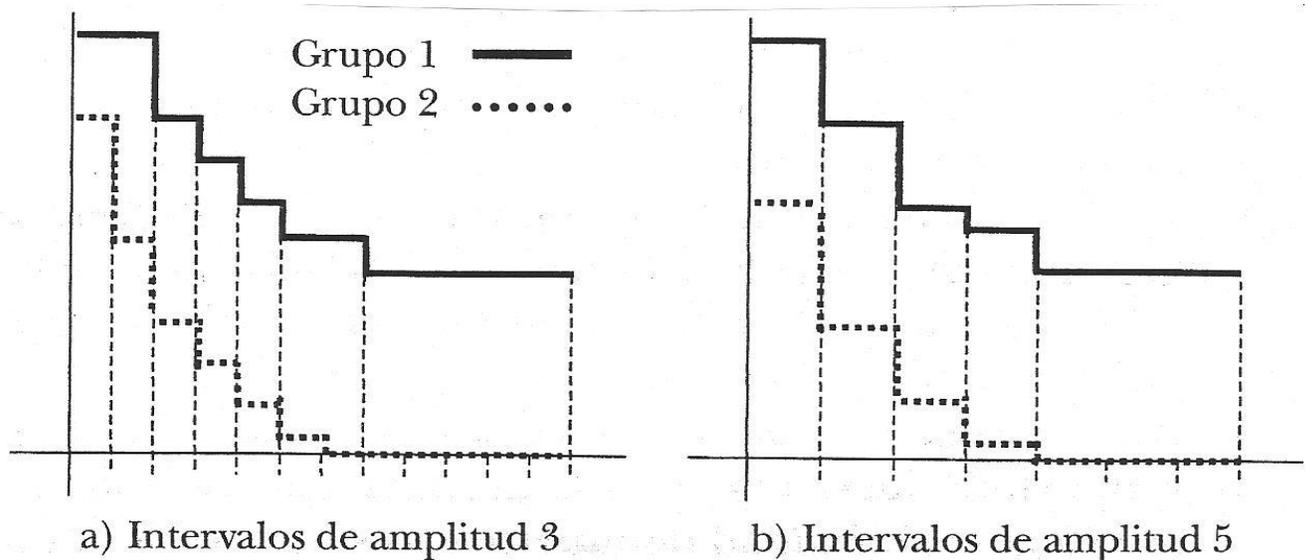
La función de supervivencia del grupo 1 a tiempo 8 semanas para estos nuevos intervalos de amplitud 5 será:

$$S_8 = 0.800 + \frac{0.686 - 0.800}{10 - 5} (8 - 5) = 0.732$$

que no es exactamente el valor hallado con intervalos de longitud 3, luego es el investigador, por su experiencia, el que debe decidir la partición en intervalos del tiempo de estudio.

Si representamos las gráficas de las funciones de supervivencia para estas dos clases de intervalos, figura 2.8, vemos que es mejor el tratamiento 6-MP que el placebo, aunque deberíamos usar un test de comparación de funciones para asegurar que esa diferencia es significativa.

Figura 2.8: Funciones de supervivencia



2.4.2 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA: PRODUCTO LÍMITE DE KAPLAN Y MEIER.

Kaplan y Meier proponen el método del producto límite con el objeto de resolver los problemas planteados por la ausencia de información en el análisis de datos para estudios de supervivencia.

El estimador de la curva de supervivencia propuesto por Kaplan y Meier se basa en el mismo principio que el actuarial: *calcular la supervivencia como producto de probabilidades condicionadas*, pero llevando la partición del tiempo de estudio en intervalos al caso extremo de considerar que cada intervalo contenga sólo la observación correspondiente a un individuo, sea ésta muerte o censura.

Si los datos observados corresponden a los tiempos $0 = t_0 < t_1 < \dots < t_n$, y se considera la partición determinada por los intervalos t_{i-1}, t_i , al que pertenece el instante de tiempo t_i pero no el t_{i-1} , y además el interior de estos intervalos está siempre libre de censuras, que sólo ocurrirán, en su caso, en un extremo.

Si llegan n_i individuos con vida al intervalo t_{i-1}, t_i , el estimador de la probabilidad de muerte en ese intervalo, condicionada a haber sobrevivido hasta entonces, será:

$$q_i = \begin{cases} i/n_i & \text{si en } t_i \text{ se produce una muerte} \\ 0 & \text{si en } t_i \text{ se produce una censura} \end{cases}$$

Los intervalos que no contienen muertes no contribuyen a la construcción de $S(t)$, ya que para ellos la estimación de la probabilidad condicionada de supervivencia en el intervalo es igual a 1. La existencia de censuras sí influye en el número de individuos expuestos al riesgo de morir al comienzo del intervalo siguiente, que se ve disminuido en una unidad.

Como indicábamos, al igual que el método actuarial, se basa en las probabilidades condicionadas pero, a diferencia de él, no establece previamente intervalos de tiempo; por el contrario, considera pequeñas unidades en relación con la frecuencia de sucesión de fallecimientos y calcula la probabilidad de supervivencia para cada una de ellas.

Si un determinado día “ i ” no hubo ningún fallecimiento: La probabilidad de sobrevivir ese día habiendo sobrevivido hasta el día “ $i-1$ ” será 1:

$$S_{i|i-1} = 1$$

Si al día “ i ” llegan n_i individuos pero fallece uno de ellos, la probabilidad de sobrevivir ese día habiendo sobrevivido el “ $i-1$ ” será:

$$S_{i|i-1} = n_i - 1 \text{ siendo esta expresión } \geq 1.$$

Este proceso se repite para todos los días y se multiplican todas estas estimaciones:

$$S_i = S_1 * S_{2|1} * S_{3|2} * S_{4|3} * \dots * S_{i|i-1}$$

De este producto se podrán suprimir los días con probabilidad de supervivencia igual a 1.

Este método tiene el inconveniente de considerar los datos incompletos como sometidos al riesgo de fallecer durante el intervalo en que finaliza su partición. Esto obliga a utilizar intervalos de tiempos pequeños en relación con el lapso de tiempo en el que van ocurriendo los fallecimientos. Hay que evitar la posibilidad de *empates* entre individuos con diferente causa de finalización.

Utilizaremos la siguiente notación:

n_i' : Número de individuos que llegan al comienzo del intervalo.

d_i : Número de fallecidos en el intervalo $t, t+1$.

p_i : Proporción de individuos que han sobrevivido al instante t .

$$p_i = \frac{n_i' - d_i}{n_i'} = 1 - \frac{d_i}{n_i'}$$

p_i es, por tanto, la probabilidad condicional de sobrevivir el i -ésimo tiempo, habiendo sobrevivido hasta el $(i-1)$ -ésimo (antes denotado por $S_{i|i-1}$)

La probabilidad de supervivencia después del instante “ t_i ” será:

$$\begin{aligned} S_{t_i} &= p_1 * p_2 * \dots * p_i \\ &= \left(1 - \frac{d_1}{n_1'}\right) \left(1 - \frac{d_2}{n_2'}\right) \dots \left(1 - \frac{d_i}{n_i'}\right) \\ &= \prod_{j=1}^i \left(1 - \frac{d_j}{n_j'}\right) = S_{t_{i-1}} \left(1 - \frac{d_i}{n_i'}\right) \\ &= S_{t_{i-1}} * p_i \end{aligned}$$

Cuando $t=0$, $S_0 = 1$; es decir, todos los individuos comienzan vivos el estudio.

2.5 PRECISIÓN DE LAS ESTIMACIONES

Los valores de \hat{q}_i, \hat{p}_i y $\hat{p} t_i$ son estimaciones sujetas a la variabilidad inherente al proceso de muestreo, por lo que deben completarse con información relativa a su precisión. Bajo determinadas hipótesis sobre los mecanismos de censura es posible, aunque complicado, deducir estimaciones de sus varianzas. Por esta razón, aunque la metodología de las tablas de vida clínicas es antigua, el estudio teórico de las propiedades estadísticas de sus estimadores es reciente y está aún por completar. En este capítulo se presentan algunas de las propiedades y resultados más utilizados. La mayor parte de estos resultados se han obtenido para el caso de muestras completas, pero se suelen generalizar y aplicar también al caso de muestras censuradas.

Fórmula de Greenwood: La estimación más empleada de la varianza de $\hat{p} t_j$ es la propuesta por Greenwood en 1926,

$$\text{Var} \left[\hat{p} t_j \right] \approx \left[\hat{p} t_j \right]^2 \sum_{i=1}^j \frac{\hat{q}_i}{\hat{p}_i n_i} = \left[\hat{p} t_j \right]^2 \sum_{i=1}^j \frac{d_i}{n_i (n_i - d_i)}$$

Esta estimación es resultado de una aproximación asintótica que comentaremos más adelante, es razonable utilizar esta aproximación cuando el valor esperado de n_j no es demasiado pequeño. La fórmula de Greenwood tiende a subestimar la varianza de $\hat{p} t_j$, especialmente en los intervalos de la cola derecha de la distribución donde el valor esperado de n_j suele ser pequeño. No obstante, en esos casos su cálculo no es adecuado ya que la distribución de $\hat{p} t_j$ suele ser muy sesgada y, en consecuencia, la varianza no es una buena medida de precisión de la estimación.

A continuación realizaremos la demostración de la Fórmula propuesta por Greenwood, para ello utilizaremos algunos supuestos sobre la distribución binomial y los polinomios de Taylor.

Justificación: La estimación actuarial de la función de supervivencia es, $\hat{p}_j = \hat{p}_j \hat{p}_{j-1} \dots \hat{p}_1$. Para obtener su varianza aproximada se aplica un procedimiento, denominado **método delta**, que consiste en calcular la varianza de la aproximación de primer orden de la función obtenida a partir de su desarrollo en serie de Taylor.

En el caso de una función de una sola variable, se tiene,

$$g(X) \approx g(\theta) + g'(\theta)(X - \theta)$$

y la varianza aproximada es,

$$\text{Var}[g(X)] \approx [g'(\theta)]^2 \text{Var}(X)$$

donde θ es un parámetro tal que, asintóticamente, $E(X - \theta) = 0$. En el caso de una función de varias variables, aplicando un procedimiento análogo, se tiene,

$$\text{Var}[g(X)] \approx \sum_{i=1}^j \sum_{k=1}^j \frac{\partial g}{\partial \theta_i} \frac{\partial g}{\partial \theta_k} \text{Cov}(X_i, X_k) \quad (2.1)$$

donde $\frac{\partial g}{\partial \theta_i}$ denota $\left. \frac{\partial g(x)}{\partial x_i} \right|_{x=\theta}$ y θ es un vector de parámetros tal que, asintóticamente,

$E(X - \theta) = 0$. Generalmente, el vector θ no es conocido, por lo que las derivadas se evalúan en su estimador $\hat{\theta}$.

Aplicando este método a la expresión \hat{p}_j , se obtiene como varianza aproximada,

$$Var\left[\hat{p}_j\right] \approx \sum_{i=1}^j \sum_{k=1}^j \frac{\partial \hat{p}_j}{\partial p_i} \frac{\partial \hat{p}_j}{\partial p_k} Cov\left(\hat{p}_i, \hat{p}_k\right)$$

Para calcular esta expresión se necesita,

1. Estimadores de p_j y q_j . Los más utilizados son:

$$\hat{p}_j = \frac{n_j - d_j}{n_j}$$

$$\hat{q}_j = \frac{d_j}{n_j}$$

En el caso sin censura, éstos son los estimadores máximo verosímiles; este resultado se obtiene utilizando el hecho que el vector d_1, d_2, \dots, d_{s+1} sigue una distribución Multinomial de parámetros n_1 , el número de individuos que inician el estudio, y $\pi_1, \pi_2, \dots, \pi_s$, la diferencia de proporciones de supervivencia en el tiempo t , donde,

$$n_1 = \sum_{i=1}^{s+1} d_i$$

$$\pi_j = p_{t_{j-1}} - p_{t_j} = p_1 \dots p_{j-1} (1 - p_j) \quad j = 1, 2, \dots, s+1$$

2. Cálculo de las derivadas parciales de la función de supervivencia respecto a cada componente del vector $p = p_1, p_2, \dots, p_s$.es decir:

$$\frac{\partial \hat{p} t_j}{\partial p_i} = \frac{\hat{P} t_j}{\hat{P}_i}$$

Demostración:

Sea $\hat{P} t_j = \hat{P}_j \hat{P}_{j-1} \dots \hat{P}_1$ el vector estimador de la función de supervivencia, si le aplicamos derivadas parciales, tendremos:

$$\begin{aligned} \frac{\partial \hat{P} t_j}{\partial \hat{P}_i} &= \frac{\partial \left(\hat{P}_j \hat{P}_{j-1} \dots \hat{P}_1 \right)}{\partial \hat{P}_i} \\ &= \left(\hat{P}_j \hat{P}_{j-1} \dots \hat{P}_{i-1} \hat{P}_{i+1} \dots \hat{P}_1 \right) \frac{\partial \hat{P}_i}{\partial \hat{P}_i} \\ &= \frac{\hat{P} t_j}{\hat{P}_i}, \text{ como se quería demostrar } \blacksquare \end{aligned}$$

3. Cálculo de las varianzas y covarianzas de los estimadores \hat{p}_i . Presentamos los resultados en el caso de muestras sin observaciones censuradas ya que su justificación en el caso general resulta complicada.

Proposición 1: Bajo la hipótesis $n_j > 0, d_j$ sigue una distribución Binomial de parámetros n_j y q_j ; entonces:

$$E \left[\hat{q}_j \right] = q_j$$

$$E \left[\hat{p}_j \right] = p_j$$

$$Var \left[\hat{q}_j \right] = Var \left[\hat{p}_j \right] = \frac{p_j q_j}{n_j}$$

$$Cov \left(\hat{q}_i, \hat{q}_j \right) = Cov \left(\hat{p}_i, \hat{p}_j \right) = 0 \quad \text{con } i < j$$

Demostración.

Como $d_j \sim B_{n_j, q_j}$ al aplicar valor esperado a la primera expresión tendremos que:

$$E \left[\hat{q}_j \right] = E \left(\frac{d_j}{n_j} \right), \text{ donde}$$

$$E \left(\frac{d_j}{n_j} \right) = \frac{1}{n_j} E d_j = \frac{1}{n_j} n_j q_j$$

entonces:

$$E \left[\hat{q}_j \right] = q_j$$

Ahora Demostraremos la segunda ecuación propuesta, utilizando siempre el supuesto que sigue una distribución binomial.

$$E\left[\hat{p}_j\right] = p_j$$

Sustituyendo al valor de \hat{P} y aplicando valor esperado tendremos:

$$\begin{aligned} E\left[\hat{p}_j\right] &= E\left[1 - \frac{d_j}{n_j}\right] \\ &= E\left[1\right] - E\left[\frac{d_j}{n_j}\right] \end{aligned}$$

Y utilizando el resultado de la primera ecuación tendremos:

$$E\left[\hat{p}_j\right] = 1 - q_j = p_j$$

En muestras con observaciones censuradas, los resultados anteriores no son ciertos aunque, asintóticamente, se verifica que $Cov\left(\hat{p}_i, \hat{p}_j\right) = 0$ y la estimación de $Var\left(\hat{p}_j\right)$ puede aproximarse mediante $\hat{p}_j \hat{q}_j / n_j$.

Sustituyendo los valores obtenidos en los puntos 1 y 2 en la expresión 2.1 de $Var\left[\hat{p}_j t_j\right]$ se obtiene la fórmula de Greenwood.

$$Var\left[\hat{p}_j t_j\right] \approx \left[\hat{p}_j t_j\right]^2 \sum_{i=1}^j \frac{\hat{q}_i}{\hat{p}_i n_i} = \left[\hat{p}_j t_j\right]^2 \sum_{i=1}^j \frac{d_i}{n_i (n_i - d_i)}$$

2.6 ESTIMACIÓN DE OTRAS FUNCIONES Y PARÁMETROS DE INTERÉS.

A partir del estimador KM se pueden obtener estimaciones de otras funciones de interés asociadas a la distribución, así como de parámetros de la misma.

Función de riesgo acumulado La función de riesgo acumulado se relaciona con $S(t)$ mediante la expresión $H(t) = -\ln S(t)$; en consecuencia, un estimador de dicha función es,

$$\hat{H} = -\ln \hat{S} \ t$$

donde $\hat{S} \ t$ es el estimador KM de la función de supervivencia.

Otro estimador posible de $H(t)$ propuesto por Nelson-Aalen, NA, es la función de riesgo acumulado empírica,

$$\hat{H}(t) = \sum_{j^{t(j)} \leq t} \frac{d_j}{n_j}$$

que acumula las contribuciones $\frac{d_j}{n_j}$ de la función de riesgo en los sucesivos instantes de

fallo t_j . La mejor estimación de la varianza de este estimador es,

$$Var \left[\hat{H}(t) \right] = \sum_{j^{t(j)} \leq t} \frac{d_j}{n_j^2}$$

A partir de este estimador de $H(t)$ es posible obtener un estimador alternativo de la función de supervivencia, denominado estimador de Nelson-Aalen, utilizando la relación,

$$\hat{S} = \exp \left[-\hat{H}(t) \right]$$

Si T es una variable continua, $\hat{H}(t)$ y $\bar{H}(t)$ son dos estimadores asintóticamente equivalentes y, salvo para valores altos de t , donde las estimaciones son más inestables, la diferencia entre ambos será, por lo general, pequeña. En realidad, $\bar{H}(t)$ es la aproximación lineal de primer orden de la función $\hat{H}(t)$. Desde el punto de vista teórico no hay argumentos para preferir un estimador al otro; el estimador $\bar{H}(t)$ tiene la ventaja de su sencillez de cálculo.

Interesa señalar la utilidad de la función de riesgo acumulado para caracterizar la distribución del tiempo de vida. Resulta más fácil analizar el comportamiento de la tasa de fallo de la distribución, representando gráficamente $\hat{H}(t)$ o $\bar{H}(t)$ frente al tiempo, que representando $\hat{S}(t)$, ya que al ser la función de riesgo la derivada de $H(t)$, se tiene que:

- Si la distribución tiene tasa de fallo constante, $H(t)$ es una función lineal en t .
- Si la distribución es monótona creciente, $H(t)$ es una función convexa.
- Si la distribución es monótona decreciente, $H(t)$ es una función cóncava.

Tiempo medio de vida Dado que la esperanza del tiempo de vida coincide con el área comprendida entre los ejes y la curva $S(t)$, un posible estimador de $E[T]$ es,

$$\hat{\mu} = \int_0^{\infty} \hat{S}(t) dt$$

Su cálculo no resulta complicado dado que $\hat{S}(t)$ es una función constante a trozos.

Cuando el máximo tiempo de supervivencia observado en la muestra, t_M , corresponde a una observación censurada, la integral anterior no puede calcularse, al no estar definido $\hat{S}(t)$ para $t > t_M$. En este caso resulta más conveniente estimar la media del tiempo de vida limitado, o restringido, al instante L , $\mu_L = E[\min T, L]$.

Tomando $L = t_M$, μ_{t_M} puede ser una buena aproximación del valor medio de T si $P(T > t_M)$ es pequeña. Se puede comprobar que,

$$\hat{\mu}_{t_M} = \int_0^{t_M} \hat{S}(t) dt$$

Un estimador de la varianza de $\hat{\mu}$ es,

$$\hat{Var}(\hat{\mu}) = \sum_r \frac{A_r^2}{n-r} \frac{1}{n-r+1}$$

donde el índice r recorre los enteros positivos tales que la observación t_r , de la muestra ordenada, no es censurada, y A_r es el área bajo la curva $\hat{S}(t)$ a la derecha de t_r . Este es un estimador sesgado por lo que es habitual -así lo hacen muchos paquetes estadísticos- corregir su sesgo multiplicándolo por un factor $\frac{n_f}{n_f - 1}$, siendo n_f el número de observaciones no censuradas de la muestra.

Percentiles y su varianza La estimación de la mediana o cualquier percentil, t_p , de la distribución es el menor tiempo de fallo observado t_j tal que, $\hat{S}[t_j] \leq 1-p$. Se puede obtener una estimación de la varianza de ese estimador aplicando el método delta al cálculo de $\text{Var}\left[\hat{S}\left(\hat{t}_p\right)\right]$. Utilizando que $S'(t) = -f(t)$ y reordenando términos se tiene,

$$\text{Var}\left[\hat{t}_p\right] = \left(\frac{1}{\hat{f}\left(\hat{t}_p\right)}\right)^2 \text{Var}\left[\hat{S}\left(\hat{t}_p\right)\right].$$

Utilizando la normalidad asintótica de \hat{t}_p y calculando con la fórmula anterior se obtiene un intervalo de confianza para t_p .

2.7 INFERENCIA SOBRE LA FUNCIÓN DE SUPERVIVENCIA.

2.7.1 ANÁLISIS GRÁFICO DE LA ADECUACIÓN DE UNA DISTRIBUCIÓN.

Antes de hacer inferencia basada en una hipótesis sobre la forma de la distribución, es necesario contrastar si esa hipótesis es adecuada a los datos observados. Para ello proponemos un procedimiento basado en la comparación de una estimación no paramétrica de $S(t)$ y la función de supervivencia del modelo elegido. Esta comparación puede hacerse gráficamente dibujando en un diagrama bidimensional $g\left[\hat{S} t\right]$ frente a $h(t)$. Las funciones g y h se eligen de acuerdo con el modelo que se quiere analizar, de forma que, si éste es adecuado, la nube de puntos se disponga, aproximadamente, según una línea recta.

Supongamos que se dispone de una muestra de tiempos de supervivencia y que se quiere analizar la hipótesis de que siguen una distribución Exponencial. Si efectivamente los datos satisfacen esta hipótesis, su función de supervivencia es,

$$S(t) = \exp -\lambda t$$

Tomando logaritmos en esta expresión y sustituyendo $S(t)$, por su estimación Kaplan – Meier, se tiene,

$$\ln\left[\hat{S} t\right] = -\lambda t$$

De esta relación se deduce que, si la hipótesis sobre el carácter Exponencial es aceptable, el gráfico de $\ln[\hat{S}_t]$ frente a t deberá ser, aproximadamente, lineal. Este análisis exploratorio nos permite además obtener una estimación preliminar del parámetro λ : la pendiente de la recta ajustada por mínimos cuadrados a la nube de puntos representada en el gráfico.

2.7.2 TEST DE LOG RANK

Consiste en comparar el número de fallecimientos observados en cada uno de los k grupos con el número de fallecimientos esperados en el caso de que la mortalidad fuera la misma en todos los grupos (hipótesis nula) La distribución de tiempos en el conjunto de la muestra (uniendo todos los grupos) permite estimar, para cada tiempo completo t_j , las proporciones de muertes que se darían bajo la hipótesis nula.

Multiplicando estas proporciones por el número de sujetos expuestos de cada grupo en el correspondiente instante se obtienen los fallecimientos esperados en cada tiempo completo. La suma de todos ellos da el número de fallecimientos esperados en cada grupo, que se comparan con los realmente observados mediante una prueba χ^2 con $k-1$ grados de libertad.

La prueba del Test Log Rank para el caso de dos grupos S, N consiste en calcular el siguiente estadístico y compararlo con la ley de χ^2 con $k-1$ grados de libertad:

En el siguiente ejemplo mostraremos todo el proceso que lleva este Test, para ello en la Tabla 7 resumiremos los cálculos necesarios para comparar las curvas de supervivencia de los pacientes con y sin trabajo estable en la que se utiliza la siguiente terminología:

O_{S_j} ; O_{N_j} Recaídas observadas en cada tiempo completo t_j , por grupo

$d_j = O_{S_j} + O_{N_j}$ Recaídas observadas en cada tiempo completo t_j , en el conjunto de la muestra (sujetos con y sin trabajo estable)

n_{S_j} ; n_{N_j} Sujetos expuestos en el instante anterior a t_j , en cada grupo.

$n_j = n_{S_j} + n_{N_j}$ Sujetos expuestos en el instante anterior a t_j , en el conjunto de la muestra (sujetos con y sin trabajo estable)

e_{S_j} ; e_{N_j} Recaídas esperadas en el instante t_j , en los grupos S y N

$Var O_{S_j} - e_{S_j}$ Varianza del número de recaídas, observadas en el instante t_j , en el grupo S , (Coincide con la varianza en el grupo N)

TABLA 7. Pacientes Con y Sin Trabajo Estable

t_j	Trabajo Estable	ESTADO	Recaídas Observadas			Sujetos Expuestos			Recaídas Esperadas		Var($O_{S_j} - e_{S_j}$) = Var($O_{N_j} - e_{N_j}$)
			O_{S_j}	O_{N_j}	d_j	n_{S_j}	n_{N_j}	n_j	e_{S_j}	e_{N_j}	
			6	N	RECUPE	0	1	1	15	17	
14	N	RECUPE									
14	N	RECUPE	0	2	2		16	31	0.9677	1.0323	0.4828
27	N	RECUPE	0	1	1	15	14	29	0.5172	0.4828	0.2497
32	N	RECUPE	0	1	1	15	13	28	0.5357	0.4643	0.2487
39	N	RECUPE	0	1	1	15	12	27	0.5556	0.4444	0.2469
48	S	RECUPE	1	0	1	15	11	26	0.5769	0.4231	0.2441
56	N	RECUPE	0	1	1	14	11	25	0.5600	0.4400	0.2464
69	N	RECUPE	0	1	1	14	10	24	0.5833	0.4167	0.2431
85	N	RECUPE									
85	N	RECUPE	0	2	2	14	9	23	1.2174	0.7826	0.4547
85	S	MUERTO									
106	N	RECUPE	0	1	1	13	7	20	0.6500	0.3500	0.2275
114	S	RECUPE	1	0	1	13	6	19	0.6842	0.3158	0.2161
132	N	RECUPE	0	1	1	12	6	18	0.6667	0.3333	0.2222
144	S	RECUPE	1	0	1	12	5	17	0.7059	0.2941	0.2076
183	S	ABS									
190	N	ABS									
218	S	RECUPE	1	0	1	10	4	14	0.7143	0.2857	0.2041
236	S	PERDIDO									
273	N	RECUPE	0	1	1	8	4	12	0.6667	0.3333	0.2222
320	N	RECUPE	0	1	1	8	3	11	0.7273	0.2727	0.1983
370	S	ABS									
392	S	RECUPE	1	0	1	7	1	9	0.7778	0.2222	0.1728
443	N	RECUPE	0	1	1	6	1	8	0.7500	0.2500	0.1875
484	S	ABS									
543	S	ABS									
638	S	ABS									
695	N	RECUPE	0	1	1	3	1	4	0.7500	0.2500	0.1875
726	S	ABS									
804	S	ABS									
908	S	ABS									
TOTAL			5	16	21				13.0755	7.9245	4.7112

La probabilidad de recaer en el instante t_j , en el supuesto de que los dos grupos tengan la misma probabilidad de recaer (hipótesis nula), se estima con el cociente d_j/n_j , que es la proporción de recaídas en el instante t_j , observadas en el conjunto de la muestra.

El número de recaídas esperadas en el instante t_j , si todos los grupos tuvieran la misma probabilidad de recaer, se obtiene aplicando la proporción de recaídas $\frac{d_j}{n_j}$, a los sujetos en riesgo en el instante anterior a t_j :

$$e_{S_j} = n_{S_j} * \frac{d_j}{n_j} ; \quad e_{N_j} = n_{N_j} * \frac{d_j}{n_j}$$

Si los cálculos están correctamente realizados para cada tiempo t_j , se verifica que:

$$d_j = e_{S_j} + e_{N_j}$$

La varianza del número de recaídas observadas en el instante t_j , para los grupos S y N viene dada por:

$$Var O_{S_j} - e_{S_j} = Var O_{N_j} - e_{N_j} = \frac{n_{S_j} n_{N_j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Por ejemplo, para el instante $t_j = 14$

$$e_{S_2} = 15 * \frac{2}{31} = 0.9677 \quad ; \quad e_{N_2} = 16 * \frac{2}{31} = 1.0323 \quad (\text{Comprobando } 0.9677 + 1.0323 = 2)$$

$$Var O_{S_2} - e_{S_2} = \frac{5 * 16 * 2 * (31 - 2)}{31^2 * (31 - 2)} = 0.4848$$

El número total de recaídas observadas y esperadas a lo largo del seguimiento y la varianza total de las recaídas en el grupo S, viene dada por:

$$\text{Recaídas Observadas: } O_S = \sum_j O_{S_j} ; \quad O_N = \sum_j O_{N_j}$$

$$\text{Recaídas Esperadas: } e_S = \sum_j e_{S_j} ; \quad e_N = \sum_j e_{N_j}$$

$$\text{Varianza: } \text{Var } O_S - e_S = \sum_j \text{Var } O_{S_j} - e_{S_j}$$

Si los cálculos de la tabla 7, han sido correctamente realizados la suma de efectivos observados debe se igual a la suma de efectivos esperados:

$$O_S + O_N = e_S + e_N \quad ; \quad 5 + 16 = 13.0755 + 7.9245 = 21$$

La prueba de Log Rank consiste en calcular el siguiente estadístico y compararlo con la ley de χ^2 , con $k-1=1$ grados de libertad. Por ejemplo a partir del grupo con trabajo S, se obtiene:

$$\chi^2 = \frac{O_S - e_S^2}{\text{Var } O_S - e_S} = \frac{5 - 13.0755^2}{4.7112} = 13.842 \quad (P = 0.0002)$$

con el grupo N se obtendría el mismo valor

$$\chi^2 = \frac{16 - 7.9245^2}{4.7113} = 13.842$$

Por lo tanto la supervivencia del grupo sin trabajo estable es peor que la del grupo con trabajo estable $\chi^2 = 13.8$; $gl = 1$; $P = 0.0002$.

Observe que las sumas de los efectivos observados, esperados y el valor χ^2 hallados coinciden exactamente con los datos de la tabla 8.

El sentido de la diferencia se obtiene al comprobar que en el grupo S , hay una menor proporción de recaídas que en el grupo N , o porque en el grupo S el número de recaídas observadas es menor a las esperadas.

Tabla 8. Resultados de comparar las curvas de supervivencia de los pacientes con y sin trabajo

Estable mediante el programa BMDPIL

	OBSERVACIONES (OBSERVADOS/ESPERADOS)	ESPERADOS	
NO	16.0	7.92	2.02
SI	5.0	13.08	0.38

<i>Test</i>	<i>Stadistic</i>	<i>Df</i>	<i>P – Valor</i>
Generalizad Savage (Mantel – Cox)	13.842	1	0.0002
Tarone – Ware	12.663	1	0.0004
Generalized Wilcoxon (Breslow)	11.617	1	0.0007
Generalized Wilcoxon (Peto – Prentice)	12.337	1	0.0004

Se puede aplicar una prueba mas conservadora utilizando al formula aproximada que da un valor ligeramente inferior (Crowley y Breslow, 1975):

$$\chi^2_S = \frac{O_S - e_S}{e_S} + \frac{O_N - e_N}{e_N}$$

$$= \frac{16 - 7.92}{7.92} + \frac{5 - 13.08}{13.08} = 13.23$$

La conclusión a partir de la formula aproximada es igual a la anterior.

2.7.3 RIESGO RELATIVO

La conclusión obtenida por la prueba Log Rank al comparar la supervivencia de los pacientes con y sin trabajo estable, tiene poco interés ya que solo permite afirmar que los sujetos sin trabajo estable tienen peor supervivencia

Una medida clásica para evaluar la magnitud de la diferencia entre las recaídas observadas a lo largo del tiempo en ambos grupos es el riesgo relativo (RR), el riesgo relativo de recaída a los sujetos sin trabajo (respecto a los sujetos con trabajo) se define como el cociente entre la incidencia de recaídas entre los sujetos sin y con trabajo estable (Elandt Jonhson y Jonhson, 1980)

una manera sencilla de evaluar el riesgo relativo es a través de la razón de riesgo (Hazard Ratio) para ello se calcula, en cada grupo, la tasa relativa de fallecimientos que es el cociente entre el número de recaídas observadas y el número de recaídas esperadas si la mortalidad fuera igual en ambos grupos (hipótesis nula)

$$Gr N = \frac{O_N}{e_N} = \frac{16}{7.9245} = 2.02 \quad ; \quad Gr S = \frac{O_S}{e_S} = \frac{5}{13.0755} = 0.38$$

estos datos indican que en el grupo N el número de recaídas observadas es el doble de las esperadas en el grupo S ocurre lo contrario ya que el número de recaídas observadas es menos de la mitad de las esperadas. El riesgo relativo se puede estimar con la razón de riesgo (Hazard Ratio) dado por la siguiente fórmula cuyo valor indica que el riesgo de recaídas se multiplica por cinco cuando el paciente no tiene trabajo estable:

$$\hat{RR} = \frac{O_N / e_N}{O_S / e_S} = \frac{2.019}{0.382} = 5.28$$

una mejor estimación del riesgo relativo la proporciona el estimador de Mantel – Hansell. Con la corrección de Tarone (1981), las formulas para calcular este estimador juntos con su error estándar y su intervalo de confianza pueden encontrarse en Breslow y Day (1987; pp. 109-110)

Existe una forma sencilla de calcular de forma aproximada el intervalo de confianza $1-\alpha$ de la razón de riesgo (Simón 1986). el logaritmo **neperiano** de la razón de riesgo viene dado aproximadamente por el cociente C :

$$\hat{EE} = \frac{1}{\sqrt{\text{Var } O_N - e_N}} = \frac{1}{\sqrt{4.7112}} = 0.4607$$

y su intervalo de confianza $1-\alpha$ viene dado por $C \pm Z_{\alpha/2} \hat{EE}(C)$:

$$\text{IC } 95\%: 1.7141 \pm 1.96 * 0.4607 = 1.7141 \pm 0.903 \quad ; \quad 0.811, 2.617$$

de manera que la razón de riesgo se estima aplicando el cambio inverso

$$\hat{RR} = e^C = e^{1.714} = 5.55 \quad \text{IC } 95\%: e^{0.811} = 2.25 \text{ a } e^{2.617} = 13.7$$

por lo tanto los paciente sin trabajo estable tienen un riesgo de recaída 5.55 veces superior al de los pacientes con trabajo estable $\text{IC } 95\% : 2.3 \text{ a } 13.7$.

CAPITULO III: APLICACIÓN DEL ANÁLISIS DE SUPERVIVENCIA.

3.1 INTRODUCCIÓN

En los dos capítulos anteriores se han descrito los conceptos y teoría sobre el análisis de supervivencia, esto nos proporciona los principios teóricos que ocuparemos para analizar la base de datos sobre los Nacimientos con Extremo Bajo Peso en el Hospital Nacional de Maternidad, para ello en este capítulo se realizará el análisis descriptivo de la mayoría de variables, además realizaremos el cruce de algunas variables que según la experiencia de algunos médicos son de mucha importancia a la hora de analizar el comportamiento de los niños con bajo peso.

La experiencia que han aportado algunos médicos del área Neonatal será base fundamental para elegir aquellas variables que pueden ser fundamentales a la hora de determinar la supervivencia y la calidad de vida que pueden tener los nacimientos que tienen las características de peso antes descritas.

El estudio de los nacimientos en este Hospital comprende cuatro áreas como son: los Datos Generales de la madre, el momento del parto y los cuidados inmediatos que tuvo el recién nacido, la etapa de Hospitalización que comprende aquellos cuidados respiratorios y los posibles problemas que puede desarrollar el recién nacido y por último la etapa de Egreso Hospitalario donde apreciaremos sobre todo la ganancia de peso y la estancia que tuvo el infante durante el proceso de Hospitalización.

Es importante mencionar que en el instante del egreso hospitalario casi el 100% de los pacientes en estudio, aun se encuentran enfermos, sin embargo, esto no indica que el paciente fallecerá en un algún período determinado, ya que el egreso hace mención a que ya se ha estabilizado para poder estar con cuidados fuera del Hospital.

Finalmente, mostraremos los estimadores que describirán la supervivencia de los niños que nacieron con las características descritas anteriormente, y que tuvieron o no algún problema o tratamiento debido a su prematurez dentro del Hospital.

3.2 DESCRIPCIÓN DE LA BASE DE DATOS Y TERMINOLOGÍA MÉDICA.

Se registraron un conjunto de variables para 127 Recién Nacidos con Extremadamente Bajo Peso al nacer en el Hospital Nacional de Maternidad, entre enero de 1999 al 31 de diciembre de 2002. Estas variables fueron divididas en cuatro etapas las cuales describiremos a continuación:

1. Datos Generales.
2. Atención Prenatal y en Sala de Partos.
3. Diagnóstico y Tratamiento.
4. Egreso Hospitalario.

3.2.1 NOMBRE DE LAS VARIABLES EN LA BASE DE DATOS.

DATOS GENERALES:

<u>Variables</u>	<u>Descripción.</u>
Nombre	Nombre de la Madre del Recién Nacido.
Exped	Número de Expediente de la madre.
Domicilio	Dirección de la madre del Recién Nacido.

PRENATAL/SALA DE PARTO:

<u>Variables</u>	<u>Descripción.</u>
C_prena	Control Prenatal.
Fec_nac	Fecha de Nacimiento.
Tipo_par	Vía del Parto (Vaginal u Operatorio).
Sexo	Sexo del Recién Nacido.
Apgar1	Apgar al Primer minuto de vida.

Apgar5	Apgar al quinto minuto de vida.
Intubac	Intubación Respiratoria.
Mas_card	Realización de Masaje Cardiaco.
Adrenal	Colocación de Adrenalina.
Oxigen	Colocación de Oxígeno.
Máscara	Colocación de Máscara respiratoria.
Peso_nac	Peso al momento de nacer.
Long_cm	Longitud al nacer.
Per_cefa	Perímetro cefálico
Edad_ges	Edad Gestacional.

DIAGNÓSTICO Y TRATAMIENTOS:

Variables

Descripción.

Oxig_cr	Necesidad de Oxígeno.
Cpap_cr	Colocación de CPAP nasal.
Vent_con	Colocación de Ventilación Convencional.
Corti_cr	Corticoides postnatal.
Hiali_cr	Enfermedad de Membrana Hialina.
Neuma_cr	Neumotórax.
Duch_art	Ductus arterioso.
Ecn_cr	ECN confirmada.
Hemor_cr	Hemorragia intracraneal.
Retinota	Retinopatía.

EGRESO:

Variables

Descripción.

Peso_egr	Peso al momento de egreso.
Fech_egr	Fecha del Egreso Hospitalario.
Con_egre	Condición de Egreso.

3.2.2 PRINCIPALES DEFINICIONES MÉDICAS UTILIZADAS.

El **APGAR** se refiere a la escala evaluativo al nacimiento, según la siguiente descripción: **Color, Tono, Llanto, Frecuencia Cardíaca y Respiración**, para calificarlo se utiliza la siguiente escala:

0 = Malo o Ausente

1 = Regular o Débil

2 = Bueno o Normal.

Una vez realizado el análisis de los cinco signos de vida al momento del nacimiento y colocado el respectivo valor según la escala antes descrita se suman dichos valores al primer y quinto minuto, si esta suma fuese mayor de 7 se tendría buenos signos de vida, pero, si dicha suma fuese menor a 7 el Recién Nacido tiene menos probabilidades de vida sobre todo si al quinto minuto no ha habido un considerable aumento en este indicador de vida. Más adelante mostraremos una tabla que describirá el comportamiento de este indicador de todos los Recién Nacidos dentro de este estudio.

El uso de algunos cuidados respiratorios es bastante normal en los nacimientos de inmaduros, por ello la experiencia ha permitido tener el conocimiento empírico de que en muchos casos el Recién Nacido que ha recibido algún cuidado respiratorio son propensos a padecer de la Enfermedad de Membrana Hialina o de Neumotórax, por ello en este caso mostraremos el cruce entre estas variables para comprobar este hecho, brindando antes la definición médica de los términos que se emplean.

CPAP Nasal: es una presión continua de vía aérea.

Ventilación Convencional: Tubo colocado a través de la boca para ayudar a que el pulmón pueda hacer su función de forma casi normal.

Corticoide: Medicamento en vena que se coloca cuando el pulmón esta inmaduro.

Membrana Hialina: Inmadures de los pulmones a causa de la inmadurez de algunas membranas.

Neumotórax: Estado de rompimiento del pulmón.

Ductus Arterioso: se refiere a una pequeña parte de las arterias pulmonares que no se cerraron al momento de nacer, pues todo tiene que cerrarse para evitar posibles padecimientos cardíacos.

ECN: Enterocolitis, es decir, infección del intestino generalmente el delgado.

La RETINOPATÍA: se refiere a la enfermedad de la retina que puede tener el Recién Nacido por la exposición a la luz al momento de nacer a causa de la inmadurez de la retina, la presencia de la retinopatía nos permite conocer que la mayoría de los Recién Nacidos Prematuros tendrán problemas de visión temprana.

3.3 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES.

Las tablas que mostraremos a continuación nos indicarán las frecuencias de algunos de los indicadores que tiene nuestro instrumento de recolección de información de los Recién Nacidos Extremadamente Prematuros.

La Tabla 1, nos indica la cantidad de madres que llevaron control prenatal en cualquier establecimiento de salud.

Tabla 1. Control Prenatal

CONTROL PRENATAL				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	68	53.5	53.5	53.5
SI	59	46.5	46.5	100.0
Total	127	100.0	100.0	

Como podemos observar solamente 59 madres llevaron control prenatal, lo que representaría el 46.5% de nuestros Recién Nacidos, este hecho marca en algún momento la prevención que se puede tener para evitar algunas complicaciones a la hora del parto.

La siguiente tabla indica el comportamiento que tiene el APGAR, quizá uno de los principales indicadores de la supervivencia del Recién Nacido Prematuro, es importante recordar que los rangos de valores que se pueden tener para este indicador son como Mínimo 1 y como Máximo 10, que por la clasificación en cuanto a peso ningún Recién Nacido lo puede obtener.

Tabla 2. APGAR al Primer y Quinto Minuto de Vida.

APGAR AL PRIMER MINUTO DE VIDA VRS. APGAR AL QUINTO MINUTO DE VIDA

APGAR AL PRIMER MINUTO DE VIDA	APGAR AL QUINTO MINUTO DE VIDA						Total
	4	5	6	7	8	9	
1		1					1
3	1		2				3
4			10	4	1		15
5			8	19	1	2	30
6			2	21	13		36
7				4	24	7	35
8					2	4	6
9						1	1
Total	1	1	22	48	41	14	127

Lo primero que observamos son los valores que aparecen en la primer columna que representan los valores del APGAR al primer minuto de vida, y que resultan de colocar la calificación respectiva según lo descrito anteriormente, mientras que los valores que aparecen en las restantes columnas representan el valor medido al quinto minuto de vida. De este cruce de datos podemos apreciar que en todos los casos hubo una mejoría de los signos de vida del primer al quinto minuto de vida, no obstante vale la pena señalar que 48 casos son los que quedaron un valor menor o igual a 7 al quinto minuto, es decir, que la mejoría no fue satisfactoria y que estos casos tiene menos probabilidad de haber sobrevivido al finalizar nuestro estudio.

Las tablas que mostraremos a continuación nos indican el tipo de reanimación que estos niños recibieron en la sala de partos, cabe aclarar que en muchos casos un niño recibió más de una forma de reanimación o que gradualmente a la evolución que iba teniendo se colocó otro tipo de reanimación para poder enviarlo a una sala de Hospitalización con el menor riesgo posible.

Tabla 3.a) Reanimación por Intubación.

REANIMACIÓN: INTUBACIÓN

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	17	13.4	13.4	13.4
SI	110	86.6	86.6	100.0
Total	127	100.0	100.0	

Como podemos apreciar 110 de los 127 niños en estudio necesitaron ser entubados al momento de nacer lo que representaría el 86.6% de nuestra población en estudio.

Tabla 3.b) Reanimación por Masaje Cardiaco.

REANIMACIÓN: MASAJE CARDIACO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	122	96.1	96.1	96.1
SI	5	3.9	3.9	100.0
Total	127	100.0	100.0	

Para el caso de la reanimación por la Técnica de Masaje Cardiaco solamente 5 de los 127 niños en estudio tuvieron la necesidad de utilizar esta técnica, lo que representa el 3.9% de la población en estudio, lo que nos indica que es poco el porcentaje de Recién Nacidos que utilizan esta técnica.

Tabla 3.c) Reanimación por Adrenalina.

REANIMACIÓN: ADRENALINA

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	122	96.1	96.1	96.1
SI	5	3.9	3.9	100.0
Total	127	100.0	100.0	

De igual manera a lo visto en la tabla 3. b) la reanimación por Adrenalina fue utilizada solamente por el 3.9% de la población, esto puede deberse a que tanto el masaje cardiaco como el uso de esteroides no son muy seguros debido a la inmadurez de la mayoría de miembros u órganos de estos niños.

Tabla 3.d) Reanimación a base de Oxígeno.

REANIMACIÓN: OXIGENO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	1	.8	.8	.8
SI	126	99.2	99.2	100.0
Total	127	100.0	100.0	

La reanimación a base de oxígeno es la más utilizada con el 99.2% del total de la población en estudio y esto puede deberse a que casi en todos los nacimientos de prematuros el mayor problema que tienen es la inmadurez de sus pulmones lo que dificulta la respiración

Tabla 3.e) Colocación de Máscara para Reanimar al Recién Nacido.

REANIMACIÓN: MÁSCARA

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	12	9.4	9.4	9.4
SI	115	90.6	90.6	100.0
Total	127	100.0	100.0	

De manera similar a lo visto en la tabla 3.a) y 3.d) la colocación de una mascarilla para ayudar a poder respirar a los recién nacidos para estabilizarlos y así transferirlos a las áreas especiales, se vuelve una de las técnicas más utilizadas en salas de parto.

Se sabe que en cuanto menos pese un Recién Nacido menor es la probabilidad de sobrevivir, es por ello que la siguiente tabla muestra el porcentaje acumulado conforme al peso de los Recién Nacidos.

Tabla 4. Peso del Recién Nacido a la hora del Nacimiento.

Peso al Nacer	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
500	1	0.8	0.8	0.8
547	2	1.6	1.6	2.4
605	1	0.8	0.8	3.1
620	1	0.8	0.8	3.9
634	1	0.8	0.8	4.7
690	1	0.8	0.8	5.5
700	4	3.1	3.1	8.7
709	1	0.8	0.8	9.4
721	2	1.6	1.6	11.0
740	1	0.8	0.8	11.8
748	1	0.8	0.8	12.6
750	3	2.4	2.4	15.0
774	1	0.8	0.8	15.7
775	1	0.8	0.8	16.5
779	5	3.9	3.9	20.5
800	9	7.1	7.1	27.6
804	1	0.8	0.8	28.3
808	6	4.7	4.7	33.1
820	2	1.6	1.6	34.6
833	2	1.6	1.6	36.2
837	3	2.4	2.4	38.6
866	3	2.4	2.4	40.9
880	4	3.1	3.1	44.1
895	9	7.1	7.1	51.2
900	14	11.0	11.0	62.2
905	1	0.8	0.8	63.0
910	1	0.8	0.8	63.8
920	16	12.6	12.6	76.4
925	1	0.8	0.8	77.2
940	1	0.8	0.8	78.0
945	1	0.8	0.8	78.7
949	7	5.5	5.5	84.3
966	1	0.8	0.8	85.0
970	1	0.8	0.8	85.8
973	1	0.8	0.8	86.6
975	1	0.8	0.8	87.4
977	2	1.6	1.6	89.0
978	10	7.9	7.9	96.9
980	3	2.4	2.4	99.2
997	1	0.8	0.8	100.0
Total	127	100	100	

Como podemos ver aquellos Recién Nacidos que están por debajo o igual a los 850 grs. representan el 38.6% de la población, lo que nos indicaría que un aproximado de 49 niños tienen mayor probabilidad de fallecer. Este hecho es reforzado cuando mostremos la cantidad de niños que fallecieron antes de finalizar nuestro estudio.

Otro indicador que es utilizado por los médicos para determinar la inmadurez de los Recién Nacidos es la edad de gestación a la hora del parto, es por ello que para brindar más apoyo al hecho que en cuanto más inmaduro es un niño mayor probabilidad de fallecer tiene, mostraremos la siguiente tabla:

Tabla 5. Edad Gestacional a la hora del parto.

Semanas de Gestación	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
20	1	.8	.8	.8
24	5	3.9	3.9	4.7
25	1	.8	.8	5.5
26	16	12.6	12.6	18.1
27	11	8.7	8.7	26.8
28	53	41.7	41.7	68.5
29	2	1.6	1.6	70.1
30	19	15.0	15.0	85.0
31	2	1.6	1.6	86.6
32	5	3.9	3.9	90.6
33	3	2.4	2.4	92.9
34	4	3.1	3.1	96.1
35	2	1.6	1.6	97.6
36	2	1.6	1.6	99.2
38	1	.8	.8	100.0
Total	127	100.0	100.0	

Como podemos apreciar 53 de los 127 niños en estudio están en el rango de las 28 semanas de gestación esto es el 68.5% del total de la población, lo cual es bastante similar a la que ocurrió en lo expuesto en la tabla 4 y como mencionan los textos médicos las semanas de gestación junto al peso es uno de los indicadores de inmadurez más utilizados, siendo el límite para la inmadurez en cuanto a semanas de gestación se refiere las 28 semanas.

Las siguientes tablas muestran los cuidados respiratorios que los Recién Nacidos tuvieron durante la estancia Hospitalaria.

Tabla 6.a) Cuidados Respiratorios con Oxígeno

CUIDADOS RESPIRATORIOS:OXIGENO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	1	.8	.8	.8
SI	126	99.2	99.2	100.0
Total	127	100.0	100.0	

De igual manera a lo ocurrido en la sala de parto la mayoría de estos niños mantuvieron oxígeno en el área hospitalaria, para este caso solamente uno de los 127 niños no tuvo necesidad de mantener oxígeno en algún momento de su internamiento.

Tabla 6.b) Cuidados Respiratorios con CPAP Nasal.

CUIDADOS RESPIRATORIOS: CPAP

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	1	.8	.8	.8
SI	126	99.2	99.2	100.0
Total	127	100.0	100.0	

Como podemos observar al igual que los cuidados con oxígeno, el CPAP Nasal es igualmente necesario para los cuidados respiratorios obteniendo 126 niños con el uso de esta técnica, recordemos que un niño puede tener más de un cuidado respiratorio y esto se debe en gran medida al desarrollo o deterioro en la salud de los infantes.

Tabla 6.c) Cuidados Respiratorios con Ventilación Convencional.

CUIDADOS RESPIRATORIOS:VENTILACIÓN CONVENCIONAL

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	8	6.3	6.3	6.3
SI	119	93.7	93.7	100.0
Total	127	100.0	100.0	

La ventilación convencional fue requerida por 119 de los 127 niños en estudio, es decir, 93.7% del total de la población en estudio, lo cual indica que esta técnica es bastante utilizada en las salas de cuidado para los recién nacidos.

Tabla 6.d) Cuidados Respiratorios con Corticoides Postnatal.

CIUDADOS RESPIRATORIOS:CORTICOIDES

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	74	58.3	58.3	58.3
SI	53	41.7	41.7	100.0
Total	127	100.0	100.0	

El uso de Corticoides es menos utilizada que las técnicas anteriores con un 41.7% del total de la población lo cual indica que solamente 53 de 127 niños en estudio recibieron este tratamiento durante su estancia en la sala de cuidados neonatales.

Ahora veremos que tipo de padecimiento cardiaco - pulmonar fue el más común para los Recién Nacidos Prematuros durante su estancia Hospitalaria, la última tabla mostrará si estos niños tuvieron algún problema hemorrágico craneal.

Tabla 7.a) Pacientes Tratados por Membrana Hialina

CUIDADOS RESPIRATORIOS:MEMBRANA HIALINA

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	19	15.0	15.0	15.0
SI	108	85.0	85.0	100.0
Total	127	100.0	100.0	

Tabla 7.b) Pacientes Tratados por Membrana Hialina

CUIDADOS RESPIRATORIOS:NEUMOTÓRAX

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	112	88.2	88.2	88.2
SI	15	11.8	11.8	100.0
Total	127	100.0	100.0	

Como podemos ver en la Tabla 7.a) 108 de los 127 pacientes desarrollaron la enfermedad de Membrana Hialina, lo cual equivale al 85.0% del total de la población. Mientras que en la Tabla 7.b) podemos apreciar que solamente 15 de los 127 pacientes mostraron el padecimiento por Neumotórax. Es decir, que el mayor problema que los niños desarrollan debido a la inmadurez de sus órganos respiratorios es la enfermedad de Membrana Hialina.

Tabla 8.a) Pacientes Tratados por Ductus Arterioso.

CUIDADOS RESPIRATORIOS:DUCTUS ARTERIOSO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	111	87.4	87.4	87.4
SI	16	12.6	12.6	100.0
Total	127	100.0	100.0	

Tabla 8.b) Pacientes Tratados por ECN.

CUIDADOS RESPIRATORIOS:ECN CONFIRMADO

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	91	71.7	71.7	71.7
SI	36	28.3	28.3	100.0
Total	127	100.0	100.0	

Como podemos apreciar en la tabla 8.a) solamente 16 de los 127 pacientes fueron tratados por Ductus Arterioso, mientras que en la tabla 8.b) 36 de los 127 niños fueron tratados por Enterocolitis, lo cual nos indica que estos dos problemas son de menor escala comparados con la Membrana Hialina o el Neumotórax.

Tabla 9. Pacientes que desarrollaron Hemorragia Intracraneal.

CUIDADOS RESPIRATORIOS

HEMORRAGIA	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
ESCASA	45	35.4	35.4	35.4
LEVE	10	7.9	7.9	43.3
MEDIA	3	2.4	2.4	45.7
CRITICA	4	3.1	3.1	48.8
NO	63	49.6	49.6	98.4
SIN EXAMEN	2	1.6	1.6	100.0
Total	127	100.0	100.0	

Como podemos apreciar solamente 63 pacientes con examen realizado no tuvieron ninguna muestra de hemorragia intracraneal, no obstante 17 pacientes tuvieron un grado de hemorragia considerable y 45 pacientes tuvieron una escasa hemorragia aunque esto no indica que el niño muera por esa causa, pero aumenta la probabilidad que estos niños tengan algún problema físico – motriz o de aprendizaje.

Tabla 10. Pacientes con Retinopatía.

RETINOPATIA

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
NO	24	18.9	18.9	18.9
SI	90	70.9	70.9	89.8
SIN EXAMEN	13	10.2	10.2	100.0
Total	127	100.0	100.0	

Como podemos apreciar 90 de los 127 pacientes mostraron algún problema de retinopatía, es decir, que estos niños tienen una gran probabilidad de desarrollar problemas visuales a temprana edad, que podrían ir desde el uso temprano de lentes hasta la ceguera temprana o desde el nacimiento. No obstante, 13 niños no fueron objeto de este estudio por lo cual de ellos no se puede tener certeza de su estado al momento del egreso hospitalario.

3.4 TABLAS DE CONTINGENCIA DE ALGUNAS VARIABLES

Ahora mostraremos una serie de tablas de contingencia con el fin de observar algunas relaciones entre los cuidados respiratorios recibidos y el desarrollo o no de alguna de las enfermedades respiratorias más comunes en Recién Nacidos, además de observar el comportamiento de algunos factores que pudieron influir en la prematurez y el peso de los Recién Nacidos en estudio.

Es importante mencionar que el haber recibido o no algún cuidado respiratorio durante su estancia dentro del Hospital no es el principal factor para asegurar la supervivencia o la calidad de vida que estos niños pueden tener, aunque nos permite tener una guía de la supervivencia de estos niños, no obstante no se puede dejar sin analizar el comportamiento que nos pueda mostrar el APGAR y los controles prenatales.

Para obtener un mejor parámetro de lo antes expuesto iniciaremos con la descripción del comportamiento que tienen la Variable Control Prenatal y el peso del Recién Nacido al momento del parto, ya que según un grupo de médicos del Hospital de Maternidad el estudio de estos niños debe de iniciar con los cuidados y seguimiento que se brinda durante el embarazo, ya que es en este momento que se obtiene por primera vez el peso aproximado que el bebe tendrá al momento que se de el nacimiento.

Tabla 11. Peso del Recién Nacido y Control Prenatal.

PESO AL NACER VRS. CONTROL PRENATAL

PESO AL NACER	CONTROL PRENATAL		Total
	NO	SI	
500		1	1
547	1	1	2
605		1	1
620		1	1
634	1		1
690		1	1
700	2	2	4
709	1		1
721	1	1	2
740	1		1
748	1		1
750	1	2	3
774		1	1
775	1		1
779	2	3	5
800	6	3	9
804	1		1
808	5	1	6
820		2	2
833		2	2
837	2	1	3
866	1	2	3
880	2	2	4
895	3	6	9
900	9	5	14
905		1	1
910		1	1
920	10	6	16
925		1	1
940	1		1
945		1	1
949	4	3	7
966	1		1
970		1	1
973	1		1
975		1	1
977	1	1	2
978	6	4	10
980	2	1	3
997	1		1
TOTAL	68	59	127

La tabla 11 nos muestra que la gran mayoría de madres no llevaron ningún control prenatal, y aun y cuando 59 pacientes si fueron objeto de algún control prenatal, otro factor a tomar en cuenta aun y cuando no se muestra en esta tabla es que en muchas ocasiones los controles se limitan a uno o dos, ocasionados generalmente por la zona geográfica donde vivían las madres.

Las tablas siguientes nos muestran la relación entre los diferentes cuidados respiratorios que tuvieron los recién nacidos y posibles enfermedades.

Tabla 12.a) Pacientes que Recibieron Oxígeno y Tuvieron Membrana Hialina.

CUIDADOS RESPIRATORIOS: OXIGENO VRS. MEMBRANA HIALINA

CUIDADOS RESPIRATORIOS: OXIGENO	CUIDADOS RESPIRATORIOS: MEMBRANA HIALINA		Total
	NO	SI	
NO		1	1
SI	19	107	126
Total	19	108	127

Esta tabla nos muestra que solamente uno de los pacientes aun y cuando no recibió oxígeno desarrollo Membrana Hialina, no obstante el dato mucho más interesante resulta en la siguiente fila, ya que si observamos la relación entre los pacientes que recibieron oxígeno y los que tuvieron problemas de Membrana Hialina estos fueron 107 pacientes contra los 19 que no desarrollaron Membrana Hialina.

Tabla 12.b) Pacientes con CPAP Nasal y Tuvieron Membrana Hialina.

CUIDADOS RESPIRATORIOS: CPAP VRS. MEMBRANA HIALINA

CPAP NASAL	CUIDADOS RESPIRATORIOS: MEMBRANA HIALINA		Total
	NO	SI	
NO		1	1
SI	19	107	126
Total	19	108	127

Como hemos visto anteriormente el uso del CPAP Nasal es igual al uso de oxígeno como método de cuidado respiratorio, es por ello que la interpretación de esta tabla es similar a la descrita en la tabla 12.a).

Tabla 12.c) Ventilación Convencional y Tuvieron Membrana Hialina.

CUIDADOS RESPIRATORIOS: VENTILACIÓN CONVENCIONAL VRS. MEMBRANA HIALINA

VENTILACIÓN CONVENCIONAL	CUIDADOS RESPIRATORIOS: MEMBRANA HIALINA		Total
	NO	SI	
NO	1	7	8
SI	18	101	119
Total	19	108	127

La tabla anterior nos muestra que 18 pacientes que estuvieron con ventilación convencional no desarrollaron Membrana Hialina, mientras que 101 estuvieron con ventilación convencional y si desarrollaron Membrana Hialina.

Tabla 12.d) Corticoides y Tuvieron Membrana Hialina.

**CIUDADOS RESPIRATORIOS:
CORTICOIDES VRS. MEMBRANA HIALINA**

CORTICOIDES	CUIDADOS RESPIRATORIOS: MEMBRANA HIALINA		Total
	NO	SI	
NO	12	62	74
SI	7	46	53
Total	19	108	127

Como hemos visto anteriormente la corticoides es de menor uso, es por ello que solamente 46 de los pacientes que tuvieron este tratamiento desarrollaron Membrana Hialina, aun así podemos observar que 62 pacientes no recibieron Corticoides pero si desarrollaron Membrana Hialina.

Las siguientes tablas nos mostraran la relación entre los cuidados respiratorios y el Neumotórax.

Tabla 13.a) Pacientes que Recibieron Oxígeno y Tuvieron Neumotórax

CUIDADOS RESPIRATORIOS: OXIGENO VRS NEUMOTÓRAX

OXIGENO	CUIDADOS RESPIRATORIOS: NEUMOTÓRAX		Total
	NO	SI	
NO	1		1
SI	111	15	126
Total	112	15	127

La tabla anterior nos muestra que solamente 15 pacientes desarrollaron Neumotórax previo cuidado con oxígeno mientras que 111 a pesar de haber estado con oxígeno no tuvieron problemas de Neumotórax.

Tabla 13.b) Pacientes que Recibieron CPAP Nasal y Tuvieron Neumotórax

CUIDADOS RESPIRATORIOS: CPAP VRS. NEUMOTÓRAX

CPAP NASAL	CUIDADOS RESPIRATORIOS: NEUMOTÓRAX		Total
	NO	SI	
NO		1	1
SI	112	14	126
Total	112	15	127

Como hemos visto en los casos de los pacientes que estuvieron con CPAP Nasal la cantidad de pacientes que tuvieron algún problema de origen respiratorio son pocos y en este caso solamente 14 de los 127 pacientes tuvieron Neumotórax luego de haber recibido el cuidado respiratorio con CPAP Nasal.

Tabla 13.c) Pacientes que Estuvieron con Ventilación Convencional y Desarrollaron Neumotórax

**CUIDADOS RESPIRATORIOS:
VENTILACIÓN CONVENCIONAL VRS. NEUMOTÓRAX**

VENTILACIÓN CONVENCIONAL	CUIDADOS RESPIRATORIOS: NEUMOTÓRAX		Total
	NO	SI	
NO	7	1	8
SI	105	14	119
Total	112	15	127

Como podemos observar solamente 14 pacientes tuvieron problemas de Neumotórax después de haber recibido ventilación convencional durante su estancia en el Hospital, mientras que 105 a pesar de haber recibido ventilación convencional no desarrollaron Neumotórax.

. Tabla 13.d) Pacientes que Recibieron Corticoides y Desarrollaron Neumotórax

CIUDADOS RESPIRATORIOS: CORTICOIDES VRS. NEUMOTÓRAX

CORTICOIDES	CUIDADOS RESPIRATORIOS: NEUMOTÓRAX		Total
	NO	SI	
NO	65	9	74
SI	47	6	53
Total	112	15	127

Como podemos observar, existen 6 paciente que recibieron Corticoides y tuvieron problemas de Neumotórax, mientras que, 47 paciente a pesar de haber recibido este tratamiento no desarrollaron Neumotórax, además como podemos apreciar solamente 9 pacientes desarrollaron Neumotórax a pesar de no haber tenido el cuidado mediante Corticoides.

3.5 ESTIMACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA.

3.5.1 TABLAS DE VIDA Y GRÁFICOS DE LAS FUNCIONES DE SUPERVIVENCIA Y DE RIESGO

Los datos que aparecen a continuación corresponden a los 127 nacimientos de niños extremadamente prematuros en el Hospital Nacional de Maternidad desde el año 1999 hasta el año 2002, para poder contar los eventos de interés y las censuras existentes se realizó una visita a cada uno de los pacientes que forman parte de este estudio en el mes de Mayo de 2003, se encontraron 49 pacientes a los que les ocurrió el evento de interés y 78 pacientes censurados, más adelante explicaremos como están divididas estas censuras.

A continuación mostraremos la Tabla de Vida, tomando para ello las siguientes variables.

estado: Ocurrencia del evento o Censura.

meses: meses vividos por ocurrencia del evento o se seguimiento por censura.

Utilizando el programa estadístico SPSS 13.5, dentro del menú analizar se elige la opción Supervivencia y luego Tablas de Mortalidad, después se divide la variable meses en intervalos de doce meses, tomando como evento de interés el fallecimiento de los niños (estado = 1) obteniendo los siguientes resultados:

Tabla de Vida

Variables de Supervivencia: MESES DE SEGUIMIENTO

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termini- nating	Propn Sur- viving	Cumul Surv at End	Proba- bility Densty	Hazard Rate
.0	127.0	22.0	116.0	22.0	.1897	.8103	.8103	.0158	.0175
12.0	83.0	51.0	57.5	26.0	.4522	.5478	.4439	.0305	.0487
24.0	6.0	5.0	3.5	1.0	.2857	.7143	.3171	.0106	.0278

La mediana del Tiempo de Supervivencia es 22.16

Intrvl Start Time	SE of Cumul Sur- viving	SE of Proba- bility Densty	SE of Hazard Rate
.0	.0364	.0030	.0037
12.0	.0568	.0046	.0091
24.0	.1146	.0090	.0274

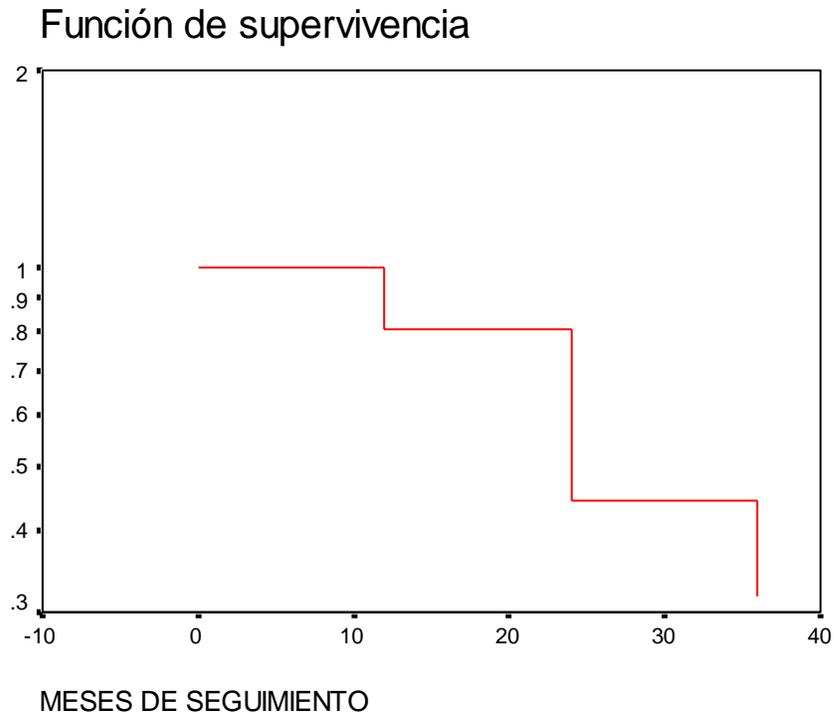
La tabla anterior nos muestra que el valor de la mediana es 22.16 meses, es decir, que aproximadamente el 50% de los niños han sobrevivido aproximadamente 22 meses y que después de este momento la supervivencia se vuelve más crítica. Es importante mencionar que por estar algunos datos censurados esta no es la simple mediana de los números que aparecen en la variable meses, por ello se ha de tener en cuenta los datos censurados, de no hacerlo así estaríamos actuando como si los datos censurados correspondieran a muertes y por tanto se produciría un sesgo hacia la izquierda.

La tercer columna de la tabla anterior nos muestra que en el primer intervalo ocurren 22 censuras, en el segundo intervalo 51 censuras y en el tercer intervalo 5, sumando entre ellas las 78 censuras mencionadas al inicio de este apartado, además podemos observar en la quinta columna que durante el primer intervalo están 22 niños a los que les ocurrió el evento de interés, en el segundo intervalo 26 eventos de interés y 1 durante el último intervalo, haciendo en total 49 niños que les ocurrió el evento de interés, recordemos que cada intervalo ha sido dividido en doce meses, que corresponden a los meses de seguimiento desde su nacimiento hasta que les ocurre el evento de interés o la censura de estos, recordemos que la búsqueda de este dato se da mediante la visita en el mes de Mayo del año 2003.

Por otra parte la proporción de supervivencia (columna 7) nos muestra que a medida van transcurriendo los años la supervivencia se vuelve menor, en este punto es importante recordar que en el estudio de supervivencia los momentos de inclusión al estudio no es el mismo y que aquellos que ingresaron primero al estudio podrían tener menos proporción de supervivencia a medida el estudio avanza que es justamente lo que muestra la columna antes señalada.

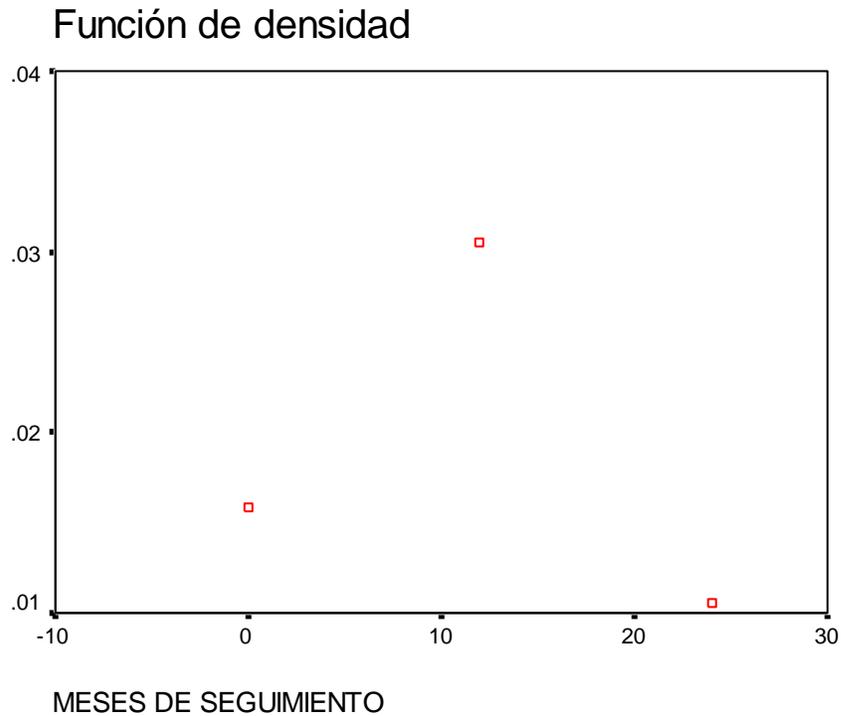
Además, debido a que el error estándar de los valores estimados es pequeño, podemos concluir que las estimaciones de las funciones de supervivencia, densidad y riesgo son buenas.

Figura1. Función de supervivencia para Recién Nacidos Extremadamente Prematuros.



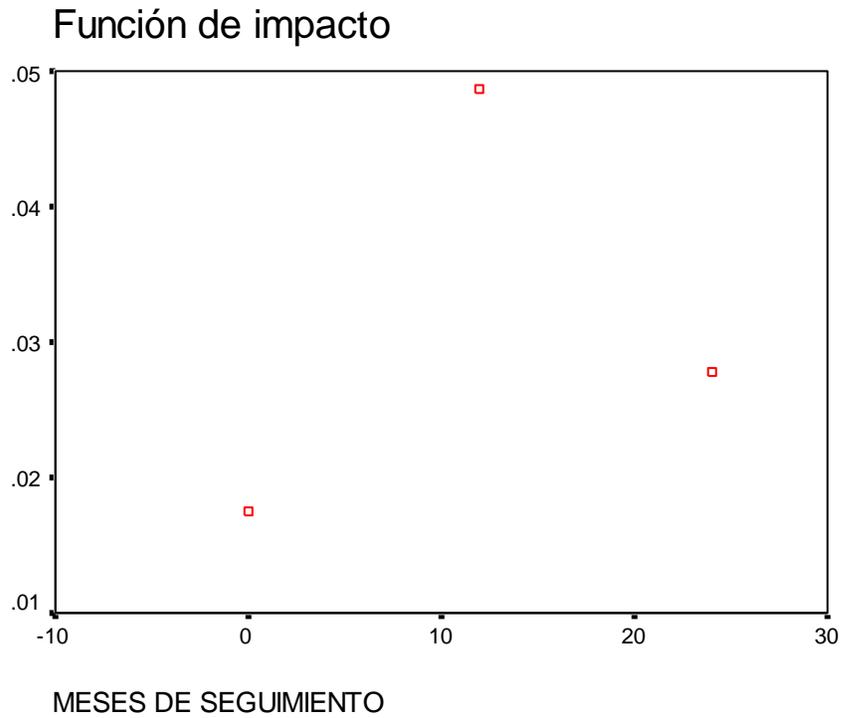
En la Figura anterior se observa que la función de supervivencia desciende progresivamente a lo largo de los 10 meses hasta caer por debajo del valor de la mediana, lo cual está acorde a lo descrito según los resultados obtenidos anteriormente.

Figura2. Función de Densidad para Recién Nacidos Extremadamente Prematuros.



La figura 2 que viene dada por los valores de la tabla en el inicio de cada intervalo, y de manera análoga a lo que apreciamos en la novena columna de la tabla de vida y ahora de forma gráfica, el intervalo de 12 a 24 meses es cuando se da el mayor riesgo de que ocurra el evento de interés, además, este intervalo contiene el valor de la mediana que como hemos dicho es de 22.16, lo cual nos da certeza que nuestro cálculos son correctos.

Figura3. Función de Riesgo para Recién Nacidos Extremadamente Prematuros.



La figura 3 corresponde a la función de riesgo (Hazard Function) que viene dada también por los valores de la tabla en el inicio de cada intervalo, y nos proporciona una idea del riesgo de muerte en cada momento para aquellos que han llegado vivos hasta ahí, y como podemos ver dicho riesgo es mayor en el intervalo de 12 a 24 meses.

3.5.2 MÉTODO DE KAPLAN – MEIER, GRÁFICOS DE SUPERVIVENCIA Y DE RIESGO.

Realizaremos un análisis similar al utilizado con el método de tablas de vida, brindando con ello mayor firmeza a las aseveraciones expuestas con los resultados de ese método. Tomando como parámetro la visita realizada a todas las madres de los Recién Nacidos en estudio, que como hemos mencionado anteriormente ocurre en el mes de Mayo del año 2003.

Se introdujo al programa estadístico SPSS las mismas variables utilizadas en la construcción de las tablas de vida (**estado y meses**), recordemos que la censura de los datos se da por abandono del estudio, por no haber encontrado a los pacientes o por que aun continúan vivos en el momento de la visita y debido a este hecho no sabemos cuando les ocurrirá el evento de interés.

Para nuestro estudio el valor “0” representará la censura y el valor de “1” la ocurrencia del evento de interés, los meses de supervivencia por ocurrencia del evento de interés o por seguimiento los mostraremos en la **tabla de la principales variables** que aparece en el anexo, donde el signo (-) representará las censuras a la izquierda; mientras que el signo (+) representará las censuras a la derecha.

Los resultados obtenidos después de evaluar las variables **estado y meses** se muestran a continuación:

Análisis de Supervivencia para la variable: MESES DE SEGUIMIENTO

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
5	CENSURADO			0	126
6	CENSURADO			0	125
6	CENSURADO			0	124
6	CENSURADO			0	123
6	CENSURADO			0	122
7	CENSURADO			0	121
7	CENSURADO			0	120
8	FALLECIDO			1	119
8	FALLECIDO			2	118
8	FALLECIDO			3	117
8	FALLECIDO			4	116
8	FALLECIDO			5	115

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
8	FALLECIDO			6	114
8	FALLECIDO	.9417	.0214	7	113
8	CENSURADO			7	112
8	CENSURADO			7	111
8	CENSURADO			7	110
8	CENSURADO			7	109
8	CENSURADO			7	108
9	FALLECIDO			8	107
9	FALLECIDO			9	106
9	FALLECIDO			10	105
9	FALLECIDO			11	104
9	FALLECIDO			12	103
9	FALLECIDO	.8894	.0290	13	102
9	CENSURADO			13	101
10	FALLECIDO			14	100
10	FALLECIDO			15	99
10	FALLECIDO			16	98
10	FALLECIDO			17	97
10	FALLECIDO			18	96
10	FALLECIDO			19	95
10	FALLECIDO	.8277	.0351	20	94
10	CENSURADO			20	93
10	CENSURADO			20	92
10	CENSURADO			20	91
10	CENSURADO			20	90
10	CENSURADO			20	89
10	CENSURADO			20	88
11	FALLECIDO			21	87
11	FALLECIDO	.8089	.0367	22	86
11	CENSURADO			22	85
11	CENSURADO			22	84
11	CENSURADO			22	83
12	FALLECIDO			23	82
12	FALLECIDO			24	81
12	FALLECIDO			25	80
12	FALLECIDO			26	79
12	FALLECIDO	.7602	.0405	27	78
12	CENSURADO			27	77
12	CENSURADO			27	76
12	CENSURADO			27	75
12	CENSURADO			27	74
12	CENSURADO			27	73
13	FALLECIDO			28	72
13	FALLECIDO			29	71
13	FALLECIDO	.7289	.0426	30	70
13	CENSURADO			30	69
13	CENSURADO			30	68
14	FALLECIDO	.7182	.0433	31	67
14	CENSURADO			31	66
14	CENSURADO			31	65
14	CENSURADO			31	64
14	CENSURADO			31	63
14	CENSURADO			31	62
15	FALLECIDO			32	61
15	FALLECIDO			33	60
15	FALLECIDO			34	59
15	FALLECIDO			35	58
15	FALLECIDO	.6603	.0470	36	57
15	CENSURADO			36	56

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
15	CENSURADO			36	55
15	CENSURADO			36	54
15	CENSURADO			36	53
15	CENSURADO			36	52
15	CENSURADO			36	51
15	CENSURADO			36	50
15	CENSURADO			36	49
15	CENSURADO			36	48
16	FALLECIDO			37	47
16	FALLECIDO	.6328	.0489	38	46
16	CENSURADO			38	45
16	CENSURADO			38	44
17	CENSURADO			38	43
17	CENSURADO			38	42
17	CENSURADO			38	41
17	CENSURADO			38	40
17	CENSURADO			38	39
17	CENSURADO			38	38
17	CENSURADO			38	37
17	CENSURADO			38	36
17	CENSURADO			38	35
18	FALLECIDO			39	34
18	FALLECIDO	.5966	.0523	40	33
18	CENSURADO			40	32
18	CENSURADO			40	31
18	CENSURADO			40	30
18	CENSURADO			40	29
18	CENSURADO			40	28
19	FALLECIDO			41	27
19	FALLECIDO	.5540	.0566	42	26
19	CENSURADO			42	25
19	CENSURADO			42	24
20	FALLECIDO			43	23
20	FALLECIDO	.5078	.0606	44	22
20	CENSURADO			44	21
20	CENSURADO			44	20
20	CENSURADO			44	19
20	CENSURADO			44	18
20	CENSURADO			44	17
20	CENSURADO			44	16
21	CENSURADO			44	15
21	CENSURADO			44	14
22	FALLECIDO			45	13
22	FALLECIDO	.4353	.0704	46	12
22	CENSURADO			46	11
22	CENSURADO			46	10
22	CENSURADO			46	9
23	FALLECIDO			47	8
23	FALLECIDO	.3386	.0815	48	7
23	CENSURADO			48	6
24	FALLECIDO	.2821	.0852	49	5
25	CENSURADO			49	4
26	CENSURADO			49	3
26	CENSURADO			49	2
28	CENSURADO			49	1
29	CENSURADO			49	0

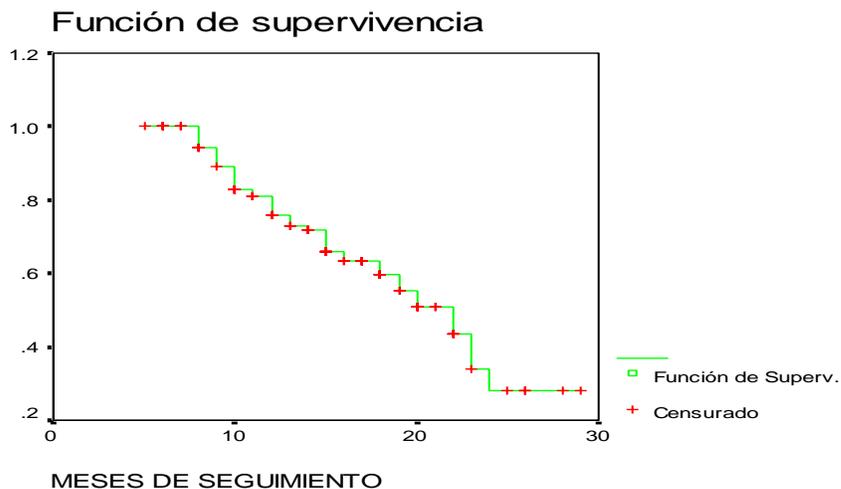
Número de Casos: 127 Censurados: 78 (61.42%) Eventos: 49

	Survival Time	Standard Error	95% Confidence Interval	
Media:	20	1	(18,	22)
Mediana:	22	2	(19,	25)

En la tercera columna podemos apreciar la función de supervivencia en aquellos tiempos en los que ocurre el evento de interés, que como hemos dicho anteriormente decae a medida va avanzando el tiempo de estudio, la cuarta columna representa el error estándar cometido al encontrar los valores de supervivencia donde ocurre el evento de interés, la quinta columna contiene la acumulación de los eventos de interés que al finalizar nos muestra como hemos dicho anteriormente 49 pacientes que fallecieron durante el período de estudio fijado.

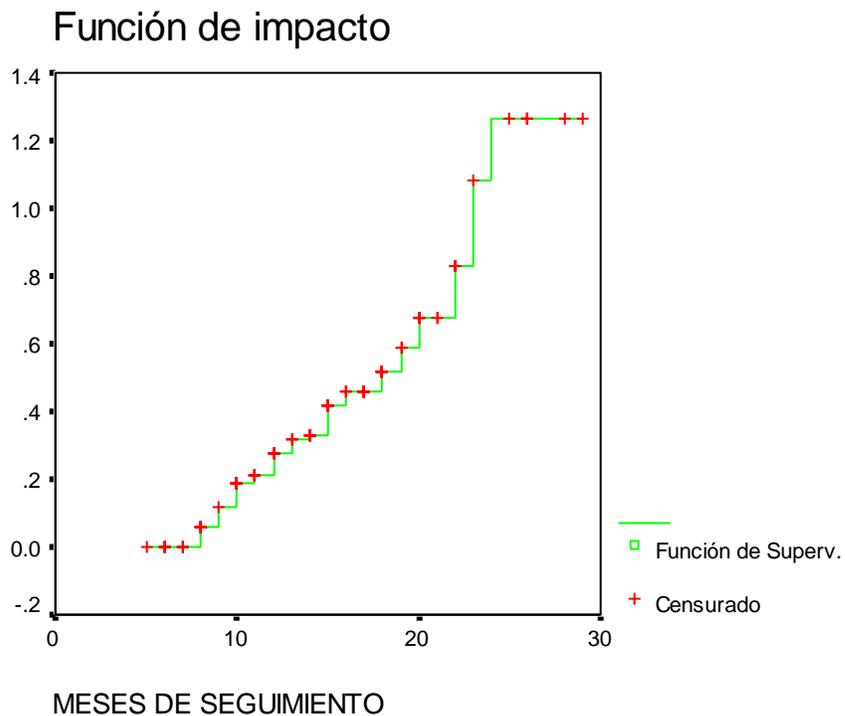
En la parte inferior podemos apreciar los valores de la media (20) y la mediana (22) del tiempo de supervivencia, que como podemos apreciar el valor entero de la mediana es el mismo que mostramos en las tablas de vida, los intervalos de confianza del 95% para la media nos indica que el límite inferior es 18 y el máximo valor es 22; mientras que para la mediana el límite inferior sería 19 y el límite superior sería 25.

Figura4. Función de Supervivencia y Censura para Recién Nacidos Extremadamente Prematuros.



La figura 4 nos brinda mayor refuerzo a lo expuesto en la gráfica de supervivencia encontrada en las tablas de vida, es decir, que a medida avanza el tiempo de estudio la supervivencia se vuelve más crítica.

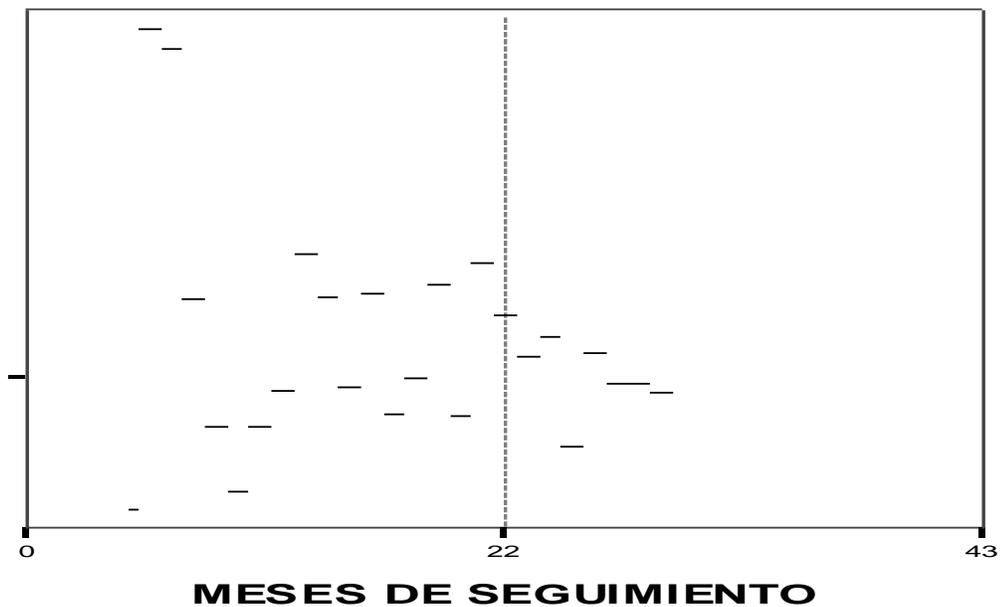
Figura5. Función de Densidad y Censura para Recién Nacidos Extremadamente Prematuros.



La figura 5 nos muestra que el riesgo de que ocurra el evento de interés aumenta a medida aumenta el tiempo de estudio, lo que esta muy acorde a lo expuesto en las gráficas y resultados mostrados anteriormente.

3.5.3 ESQUEMA DE CENSURA

A continuación se muestra el esquema de censura, tomando como valor de referencia la mediana encontrada anteriormente, los meses de seguimiento se muestran en la **tabla de las principales variables** que aparece en el anexo, en la cual veremos que la mayor cantidad de meses de supervivencia es de 29, esto debido a que recordemos que cada paciente ingresa en diferente fecha, siendo la principal cualidad que tiene el Análisis de Supervivencia respecto a las otras ramas de la estadística.



Las líneas a la izquierda de la vertical punteada representan las censuras por la izquierda, es decir, son aquellos pacientes que al momento de la visita no fueron encontrados o que abandonaron el estudio, de estos conocemos en forma aproximada cuantos meses habían sobrevivido, pero al no tener la certeza de la ocurrencia del evento de interés, representamos ese valor acompañado con el signo (-); mientras que las líneas a la derecha representan las censuras por la derecha, es decir, son aquellos pacientes que al momento de la visita no les ha ocurrido el evento de interés y por consiguiente no sabemos cuando les ocurrirá, en este caso el valor de supervivencia está acompañado por el signo (+), como se muestra en la tabla con las principales variables del estudio que aparece en el anexo.

3.6 CONCLUSIONES

- La mortalidad de los recién nacidos extremadamente prematuros es de 45.00% en el primer año de estudio, 25.00% en el segundo año, 20.00% durante el tercer año y de 10.00% durante el último año de estudio.
- La probabilidad de supervivencia disminuye a medida se avanza en el estudio, causando que el riesgo de fallecer aumente en cada período.
- Debido a que el valor del error estándar es pequeño podemos concluir que los métodos utilizados son correctos y que el comportamiento mostrado durante el estudio es apegado a la realidad de este tipo de nacimientos.
- El mayor problema de los nacimientos extremadamente prematuros es el desarrollo de la enfermedad de Membrana Hialina, seguidamente por el Neumotórax.
- La calidad de vida de los que al finalizar el estudio no les ha ocurrido el evento de interés (muerte), estará marcada por problemas de *Retinopatía*, disminución de las funciones cerebrales, debidas a las *Hemorragias Intracraneales* o *Bronco Pulmonares*, ocasionadas por los problemas mostrados durante el período de hospitalización.
- Las causas probables de que ocurra el evento de interés se deben primero a que el parto se da antes de las 27 semanas de gestación, además, que el peso de estos niños no sea mayor a los 800 grs. y que desarrollen algún tipo de enfermedad o padecimiento debido a la inmadurez de algunos de sus órganos.
- El aumento en los controles prenatales podría ser un factor determinante para poder evitar los nacimientos prematuros o mejorar la calidad de vida de los nacidos de forma extremadamente prematura.

ANEXOS

PRINCIPALES VARIABLES DE BASE DE DATOS.

Fecha Nacim.	Fecha Egreso	Peso Nacim.	APGAR1	APGAR5	Sem. Gesta.	Estado	Superv. Meses
14.02.99	27.03.99	866	4	6	32	1	8
19.02.99	08.04.99	940	5	7	29	1	9
08.04.99	03.07.99	895	4	7	30	1	9
22.05.99	12.11.99	800	6	8	25	0	10 -
25.05.99	10.08.99	837	4	6	26	1	10
25.05.99	25.08.99	800	6	7	30	0	10 -
06.06.99	14.08.99	837	5	6	28	1	10
11.06.99	07.09.99	895	6	7	26	0	9 -
26.08.99	18.12.99	800	6	8	27	0	11 -
04.09.99	22.11.99	920	7	8	28	0	11 -
09.09.99	18.11.99	700	7	8	28	0	8 -
15.09.99	17.11.99	900	7	8	28	0	10 -
20.09.99	08.12.99	920	5	9	27	0	15 -
24.09.99	04.03.00	750	5	7	27	1	12
18.10.99	18.11.99	978	7	9	20	0	17 -
02.11.99	24.12.99	779	6	8	28	0	10 -
16.11.99	10.01.00	895	8	9	31	0	15 -
16.11.99	06.04.00	866	4	6	26	1	10
23.11.99	03.03.00	880	7	9	26	0	14 -
09.12.99	29.11.00	547	7	8	28	0	12 -
18.12.99	24.02.00	978	6	7	28	0	15 -
16.01.00	20.04.00	775	4	6	32	1	10
16.01.00	06.05.00	800	3	6	28	1	11
17.01.00	19.06.00	808	6	8	32	0	18 -
25.01.00	15.02.00	949	7	8	24	0	20 -
04.02.00	08.03.00	500	6	8	28	0	15 -
06.02.00	05.05.00	978	6	7	30	0	17 -
21.02.00	22.06.00	966	6	8	28	0	17 -
28.02.00	22.06.00	779	4	8	28	0	12 -
15.03.00	13.06.00	978	7	7	28	0	13 -
25.03.00	16.06.00	709	4	7	27	1	12
29.03.00	22.05.00	920	4	6	30	1	12
02.04.00	26.06.00	634	4	6	26	1	14
20.04.00	15.06.00	978	7	8	28	0	20 -
01.05.00	10.06.00	949	7	9	38	0	20 -
04.05.00	03.07.00	900	5	7	28	1	15
05.05.00	29.07.00	895	7	8	33	0	18 -
23.05.00	25.08.00	700	7	8	24	0	15 -
07.06.00	21.07.00	920	5	7	32	1	15
20.06.00	04.10.00	977	3	4	30	1	8
06.07.00	23.10.00	895	7	8	28	0	16 -
10.08.00	08.11.00	808	8	8	28	0	17 -
21.08.00	28.12.00	978	6	7	26	0	17 -

Fecha Nacim.	Fecha Egreso	Peso Nacim.	APGAR1	APGAR5	Sem. Gesta.	Estado	Superv. Meses
24.08.00	07.11.00	900	4	6	31	1	9
20.09.00	20.11.00	920	5	9	27	0	25 -
23.09.00	07.01.01	774	5	6	27	1	8
08.10.00	20.12.00	833	3	6	28	1	8
28.10.00	05.02.01	779	7	7	27	0	12 -
11.11.00	29.01.01	895	6	8	28	0	15 -
13.11.00	22.03.01	804	5	7	30	1	10
19.11.00	05.02.01	920	6	7	33	0	15 -
21.11.00	13.02.01	920	5	7	29	1	10
31.12.00	25.02.01	978	8	9	35	0	29 +
13.01.01	12.03.01	920	4	7	34	1	23
17.01.01	13.03.01	949	7	8	30	0	28 +
23.01.01	10.06.01	748	6	6	26	1	23
05.02.01	28.04.01	920	4	6	28	1	22
09.02.01	03.05.01	978	5	6	30	1	20
10.02.01	15.06.01	750	5	7	28	1	19
14.02.01	28.05.01	700	7	9	35	0	17 -
20.02.01	08.05.01	800	5	7	30	1	18
28.02.01	07.05.01	900	6	8	30	0	20 -
05.03.01	10.04.01	997	7	9	34	0	26 +
07.03.01	29.05.01	833	6	7	28	0	18 -
21.03.01	06.07.01	977	5	7	28	1	20
31.03.01	08.06.01	700	9	9	27	0	26 +
18.04.01	07.08.01	837	5	7	28	1	24
23.05.01	05.08.01	800	6	7	30	0	20 -
30.05.01	08.08.01	973	5	6	30	1	22
13.06.01	26.10.01	900	5	8	36	0	20 -
20.06.01	22.08.01	949	6	7	30	0	18 -
24.06.01	05.11.01	920	7	8	28	0	23 +
02.07.01	10.10.01	949	5	7	28	1	19
04.07.01	07.09.01	895	7	7	28	0	22 +
04.07.01	22.09.01	895	7	8	28	0	22 +
05.07.01	22.08.01	920	7	7	26	0	22 +
04.08.01	03.10.01	808	6	7	28	0	18 -
10.08.01	12.10.01	721	4	6	28	1	18
10.08.01	31.10.01	900	6	7	28	0	18 -
15.08.01	15.08.01	808	6	7	28	0	17 -
21.08.01	12.10.01	900	7	8	28	0	21 +
16.09.01	21.11.01	920	7	8	28	0	21 +
25.10.01	31.12.01	980	5	7	30	1	16
30.10.01	09.01.02	978	7	8	28	0	19 +
30.10.01	27.12.01	910	7	8	30	0	19 +
13.11.01	24.02.02	880	5	7	28	1	16

Fecha Nacim.	Fecha Egreso	Peso Nacim.	APGAR1	APGAR5	Sem. Gesta.	Estado	Superv. Meses
20.11.01	13.02.02	750	7	8	28	0	16 -
03.12.01	17.03.02	900	4	7	27	1	15
13.12.01	20.02.02	920	6	8	28	0	17 +
21.12.01	16.03.02	920	5	7	26	1	15
27.12.01	20.02.02	808	5	6	27	1	15
01.01.02	05.03.02	820	7	8	28	0	17 +
04.01.02	19.04.02	820	1	5	28	1	13
24.01.02	20.03.02	721	6	7	28	1	13
25.01.02	26.04.02	800	5	7	26	1	13
01.02.02	14.05.02	800	6	6	26	1	12
11.02.02	12.07.02	980	7	8	28	0	15 +
12.02.02	10.09.02	620	7	8	24	0	15 +
14.02.02	22.05.02	920	6	7	28	0	15 +
14.02.02	22.05.02	690	6	7	24	1	12
03.03.02	27.05.02	895	7	8	30	0	14 +
12.03.02	20.05.02	900	7	9	28	0	14 +
22.03.02	22.10.02	547	7	8	26	0	14 +
31.03.02	24.07.02	800	8	9	34	0	14 +
09.04.02	06.06.02	900	5	7	32	1	11
14.04.02	16.07.02	808	6	7	28	0	13 +
04.05.02	30.06.02	949	5	7	28	1	9
10.05.02	25.07.02	900	5	6	28	1	9
15.05.02	07.08.02	978	6	8	30	0	12 +
17.05.02	11.07.02	866	6	7	30	0	12 +
03.06.02	16.08.02	920	4	6	30	1	8
18.06.02	19.08.02	900	6	8	28	0	11 +
30.06.02	21.10.02	779	5	6	26	1	10
03.07.02	18.09.02	980	8	8	28	0	10 +
03.07.02	09.09.02	949	6	7	28	0	10 +
13.07.02	05.10.02	779	5	7	27	1	8
30.07.02	18.12.02	605	5	6	26	1	9
27.08.02	24.12.02	900	5	7	26	1	8
06.09.02	22.01.03	880	6	8	28	0	8 +
06.09.02	11.01.03	880	6	8	28	0	8 +
09.09.02	22.12.02	905	7	9	33	0	8 +
18.10.02	20.12.02	900	7	8	28	0	7 +
18.10.02	28.10.02	740	8	9	36	0	7 +
10.11.02	18.02.03	970	6	7	26	0	6 +
12.11.02	13.01.03	945	7	8	34	0	6 +
23.11.02	07.02.03	975	6	7	28	0	6 +
17.12.02	05.04.03	925	6	7	24	0	6 +