UNA REVISIÓN DE ALGUNAS METODOLOGÍAS ESTADÍSTICAS EN EL ANÁLISIS DE SUPERVIVENCIA

CARMEN ASTRID VIVEROS LEDEZMA
JUAN MANUEL FUELANTALA

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN
DEPARTAMENTO DE MATEMÁTICAS
POPAYÁN
2009

UNA REVISIÓN DE ALGUNAS METODOLOGÍAS ESTADÍSTICAS EN EL ANÁLISIS DE SUPERVIVENCIA

Seminario de grado presentado como requisito
parcial para optar el titulo de:

Licenciados en Educación con Especialidad en Matemáticas a

CARMEN ASTRID VIVEROS LEDEZMA
JUAN MANUEL FUELANTALA

Director:
Mg. EDWIN RENGIFO

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN
DEPARTAMENTO DE MATEMÁTICAS
POPAYÁN
2009

Dedicamos este trabajo a quienes nos
apoyaron incondicionalmente:
nuestros padres, hijos, familiares y amigos.

Nota de aceptación:

Director:

Mg. Edwin Rengifo

Comité evaluador:

Mg. Carlos Julio Restrepo

Mg. Yilton Riascos Forero

Fecha de sustentación:

Agradecimientos

Los autores agradecemos los comentarios positivos y las valiosas observaciones que recibimos de:

Edwin Rengifo, profesor del departamento de Matemáticas y director de este trabajo por sus valiosos aportes, y constante empeño y dedicación,

Luis Eduardo Montoya, profesor del departamento de Matemáticas por su dedicación, colaboración y sugerencias en la realización del trabajo,

Carlos Julio Restrepo, profesor del departamento de matemáticas y miembro del comité evaluador, por sus valiosas sugerencias para ser posible este trabajo,

Yilton Riascos Forero, profesor del departamento de matemáticas y del comité evaluador, por su colaboración y sugerencias para hacer posible este trabajo,

A nuestras familias, por su apoyo, paciencia y colaboración,

A la Universidad del Cauca, el lugar que nos permitió crecer como ciudadanos y profesionales.

Contenido

Resumen

En este documento se presenta el informe del seminario de grado "Una revisión de algunas metodologías estadísticas no-paramétricas en el análisis de supervivencia", realizado dentro del grupo de estudio y desarrollo investigativo en matemática aplicada en la línea de Estadística y Probabilidad del Departamento de Matemáticas.

En el capitulo 1, se dan a conocer los conceptos básicos del análisis de supervivencia como: la función de supervivencia, la fuerza de mortalidad, y la esperanza de vida a partir de algunas funciones de la teoría de la probabilidad. Se consideran las probabilidades condicionales de supervivencia para estimar probabilidades de un evento en un determinado período de tiempo.

En el capitulo 2, se establecen las características propias de algunas metodologías estadísticas, se describen los métodos estadísticos no-paramétricos en el análisis de supervivencia como: el método de la tabla de vida, el método de Kaplan-Meier y el método de Log-Rank. Se destaca la estimación de probabilidades condicionales para obtener la función de supervivencia en cada método. Ejercicios de aplicación.

Introducción

El análisis de supervivencia tiene sus origenes en la Ingenería, disciplina en la cual dicho análisis esta encaminado a estudiar el tiempo de duración y fiabilidad de los diferentes elementos que forman una máquina. Hoy en día este concepto es aplicable en diferentes ramas de estudio, incluso para analizar la supervivencia humana, de animales o de modelación de comportamientos de objetos, empresas o bienes sujetos a ciertos procesos. El análisis de supervivencia se interpreta como una metodología estadística para analizar datos para los cuales la variable respuesta de interés esta relacionada a la medición del tiempo que transcurre hasta la valoración de un determinado evento o suceso.

Un evento que se defina evaluar en una muestra de unidades donde se presupone una distribución particular de la variable, o bien sus hipótesis especifican parámetros o distribuciones, con frecuencia se usan los métodos estadísticos paramétricos; sin embargo, cuando no se presuponen que los datos se ajusten a una distribución concreta, las variables no necesariamente deben estar medidas en un nivel por intervalos o razón, pueden analizarse datos nominales u ordinales, por lo que en realidad no tenemos parámetros a estimar sino distribuciones que comparar. En este caso, se consideran los métodos estadísticos no-parámetricos. Estos métodos se adaptan lo suficientemente bien al análisis de tiempos de supervivencia; ya que una de las principales razones estriba en que las distribuciones para los tiempos de supervivencia presentan características especiales que no a costumbran a recogerse con los modelos probabilísticos usuales. Otra ventaja, es que se tiene en cuenta el carácter secuencial de los datos y los ajusta de forma que cada unidad sólo contribuye al estudio mientrás está bajo observación.

Uno de los propósitos de este documento es el dar a conocer algunos conceptos elementales en el análisis de supervivencia para la descripción y resumen de los tiempos de supervivencia, usualmente a partir de la estimación e interpretación de la función de supervivencia y la fuerza de mortalidad en donde se considera una distribución empírica de los datos, por lo cual los resultados estadísticos se derivan de las mismas; puesto que no se presupone una distribución concreta de la variable que representa la supervivencia, ni tampoco cumple con una normalidad, se utilizan en el análisis algunos métodos estadísticos no-paramétricos como: el método de la tabla de vida, el método de Kaplan-Meier, y para la comparación de las distribuciones de los tiempos de supervivencia correspondientes a dos poblaciones diferentes y el establecimiento y comprensión de la relación que pueda existir entre los tiempos de supervivencia y las variables explicativas se recurre al método de Log-Rank o prueba del intervalo logarítmico.

En la bibliografía consultada referente a los métodos estadísticos no-paramétricos en el análisis de supervivencia, se encontró que comprenden conceptos complejos que exijen para su interpretación tener conocimientos básicos en el tema. El aporte presentado en este documento tiene como objetivo principal el brindar a estudiantes universitarios y profesionales en el análisis de supervivencia una introducción a los conceptos avanzados de la estadística que en él se manejan, bajo el soporte de algunas nociones elementales de la teoría de la probabilidad y de las Matemáticas.

Capítulo 1

Conceptos básicos del análisis de supervivencia.

1.1. Introducción

Los estudios de supervivencia[1], se basan en investigaciones en las cuales una muestra de unidades es observada por un período de tiempo llamado Período de observación o tiempo de seguimiento durante el cual se espera que se produzca el evento[2] o acontecimiento de interés del estudio. El tiempo que transcurre desde el momento de entrada de las unidades (por primera vez) al estudio, hasta que se alcanza el objetivo determinado, es llamado Tiempo de Supervivencia. Deben considerarse en los estudios, excepciones que los caracterizan, como la presencia de observaciones incompletas de un período, llamadas Censuras que surgen por diversas razones: algunas de las unidades se pierden o desaparecen del estudio, o en la fecha de cierre del estudio aún no se ha producido el evento de interés, o algunas unidades experimentan un evento diferente al esperado; pero que imposibilita el seguimiento. La información (tiempo) que aportan las observaciones incompletas o censuras de las unidades que no completaron el período de seguimiento, es llamado Tiempo Censurado. Pueden considerarse tres tipos de censuras: Censura por la derecha, el tiempo de seguimiento finalizó antes de que el evento de interés ocurriese. Censura por la izquierda, el evento ocurrió antes de que la unidad de estudio fuese observado. Censura en intervalos, el evento ocurrió durante un

[1] La medición del tiempo que transcurre hasta que sucede un evento o acontecimiento de interés. Castelazo A., Luis. Análisis de sobrevivencia pacientes de ginecología, reporte No. 4, hospital de ginecobstetra, México. Junio 2001, pág. 13.

[2] Un evento es un acontecimiento, una eventualidad, hecho imprevisto, o cosa que puede suceder. Microsoft Student con Encarta Premium 2008.

intervalo de tiempo; pero no se sabe exactamente en que instante de tiempo.

Las censuras mencionadas pueden ilustrarse a través del ejemplo expuesto por Cesar Peréz en su libro "Técnicas estadísticas con SPSS", capitulo 12, pág. 428.

Considérese el análisis del tiempo que transcurre entre el diagnóstico de un tipo de cáncer avanzado (sin posibilidad de cura) en un grupo de pacientes y la muerte de los mismos. Los pacientes son observados cada seis meses, empezando justo en el momento en que se les diagnosticó el cáncer.

- Caso 1 (Censura por la derecha): Supóngase que a todos los pacientes se les diagnosticó el cáncer el mismo día. Transcurridos los seis meses, es posible que algunos pacientes hayan muerto y otros no. Para los sobrevivientes la duración o tiempo de supervivencia es mayor al período observado de seis meses.

- Caso 2 (Censura por la izquierda): Supongamos que estamos interesados en la supervivencia de un grupo de pacientes con síntomas de un determinado tipo de cáncer, donde algunos de ellos ya se les ha diagnosticado un cierto tipo de cáncer y algunos de los restantes posiblemente mueran antes de ser diagnosticados.

- Caso 3 (censura de intervalo): Supongamos que algunos de los pacientes a quienes ya se les diagnosticó un tipo de cáncer, y que han sobrevivido a seis meses de observación, mueren un año después. Esto quiere decir, que el evento de interés, la muerte, ocurrió entre seis meses y un año.

Nótese que, los datos de supervivencia son un conjunto de observaciones cada una representando un episodio de tiempo, al final del cual ocurre o no el evento estudiado. Al análisis de este tipo de datos se le llama Análisis de Supervivencia, y es una metodología estadística que permite estudiar mediante una función de distribución para una variable aleatoria que representa el tiempo de supervivencia, probabilidades de supervivencia y de riesgo[3]. Hay que destacar las diferencias que existen entre éstas dos probabilidades. La probabilidad de supervivencia o "la probabilidad de la no ocurrencia del evento de interés" es posible mediante función de supervivencia, mientrás la probabilidad de riesgo o "la probabilidad de ocurrencia del evento" se obtiene de la función de riesgo o fuerza de mortalidad.

[3]El riesgo se puede interpretar como la incertidumbre que existe de que un suceso o evento ocurra.

Antes de dar a conocer los conceptos básicos en el análisis de supervivencia que se tratarán en este documento, muy brevemente se presentan los conceptos necesarios de la teoría de la probabilidad para que el lector pueda incurrir en el estudio del análisis de supervivencia.

1.2. Algunos conceptos de la Teoría de la Probabilidad.

Brevemente se presentan los conceptos necesarios de la teoría de la probabilidad para comprender procedimientos matemáticos y estadísticos que se hacen a lo largo del documento; aunque es necesario que el lector se apoye en otras fuentes bibliográficas.

Definición 1. ***(Espacio Muestral)***
Al conjunto finito formado por todos los posibles resultados de un experimento, se le llama Espacio Muestral o Espacio Elemental de eventos. Cada uno de los posibles resultados se llama evento elemental o punto muestral.

Sea Ω el espacio muestral no ordenado,

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$$

donde, los ω_i son los eventos elementales o puntos muestrales.

Definición 2. ***(Espacio de Probabilidad)***
Consideremos un experimento no-determinístico[4] *cuyo espacio muestral es Ω, y sea $\mathcal{A}$ una álgebra de subconjuntos asociados a Ω. Si se le asigna a cada punto muestral $\omega_i \in \Omega$; $i = 1, 2, \ldots, N$, un peso el cual se representa por $p(\omega_i)$, y es llamado La Probabilidad de Ocurrencia del resultado ω_i; y suponemos que se cumplen las siguientes propiedades:*

1. $0 \leq p(\omega_i) \leq 1$ *(Propiedad de positividad)*

2. $p(\omega_1) + p(\omega_2 + \ldots + p(\omega_N) = 1$ *(Propiedad de normalización)*

[4]Esto es, cuando el espacio muestral contiene más de un resultado posible.

A partir de las probabilidades de $p(\omega_i)$ y de los resultados (ω_i), se define La Probabilidad de cualquier evento $A \in \mathcal{A}$, denotado $P(A)$ tal que:

$$P(A) = \sum_{\{i:\ \omega_i \in A\}} p(\omega_i) \tag{1.1}$$

A la terna $(\Omega, \mathcal{A}, P)$ donde $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ es el espacio muestral, $\mathcal{A}$ es una álgebra de subconjuntos asociada a Ω y $P : \mathcal{A} \longrightarrow [0,1]$ la función definida por la expresión (1.1); define un Modelo de Probabilidad o Espacio de Probabilidad al experimento no-determinístico subyacente.

Consideremos el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y dos eventos cualesquiera $A, B \in \mathcal{A}$, se cumplen las siguientes propiedades:

- $P(\emptyset) = 0$
- $P(\Omega) = 1$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si $A \subseteq B$, entonces, $P(A) \leq P(B)$
- Si $A \subseteq B$, entonces, $P(B - A) = P(B) - P(A)$
- Si $A \cap B = \emptyset$, entonces, $P(A + B) = P(A) - P(B)$

Nota: *Si se considera el espacio muestral Ω un conjunto finito o infinito numerable, se entenderá que el álgebra de eventos es por defecto el conjunto de partes de Ω, $\wp(\Omega)$. Si llegase hacer Ω un conjunto infinito no numerable como $\mathbb{R}$, $\mathbb{R}^+$ o subconjuntos de éstos en forma de intervalos, se entenderá que el álgebra de subconjuntos es la formada por todos los intervalos abiertos, cerrados, semi- abiertos, semi- cerrados, y sus uniones finitas incluyendo en particular a los puntos.*

Definición 3. ***(Variable Aleatoria)***
Sea $(\Omega, \Im, P)$ un espacio de probabilidad inicial. Consideremos $(\Omega, \Im)$ y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dos espacios medibles, toda transformación medible $X : \Omega \longrightarrow \mathbb{R}$ se llama variable

aleatoria[5] *puesto que si* $\mathcal{B}(\mathbb{R})$ *es la colección de todos los subconjuntos de* $\mathbb{R}$*, y sea* $B \in \mathcal{B}(\mathbb{R})$ *cualquier subconjunto, la función* X *le asocia a todo punto muestral* $\omega \in \Omega$ *un número real* $x \in \mathbb{R}$*, esto es* $X(\omega) = x \in \mathbb{R}$ *y debe satisfacer la condición de que* $X^{-1}(B)$ *es un elemento del álgebra* $\Im$*. Esto se expresa de la siguiente forma:*

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \Im; \ \forall B \in \mathcal{B}(\mathbb{R})$$

Nota: En la teoría de probabilidad no importa tanto la forma funcional de la variable X, sino más bien, la probabilidad de que un valor observado de la variable aleatoria X pertenezca a un conjunto dado $B \in \mathcal{B}(\mathbb{R})$, representado por $P_X(B)$. Así, una variable aleatoria es un mecanismo que da lugar a un experimento no-determinístico con resultados numéricos. La variable aleatoria induce un nuevo espacio de probabilidad: $(\Omega, \Im, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$

Definición 4. ***(Función de Probabilidad)***
Dado el espacio muestral Ω*, y* X *una variable aleatoria. La función de probabilidad de* X *denotada por* P_X*, tal que* $P_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$ *es una función definida para todo conjunto boreliano* $B \in \mathcal{B}(\mathbb{R})$ *cuyo valor* $P_X(B)$ *es la Probabilidad de que* X *esté en* B*. Esta probabilidad se define por:*

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}); \ \forall B \in \mathcal{B}(\mathbb{R}) \tag{1.2}$$

$P_X(B)$ *puede expresarse como* $P(X \in B)$*, y significa que la variable aleatoria* X *tendrá un valor observado perteneciente al conjunto* B *si, sólo si, el valor observado* ω *del experimento no-determinístico es tal que* $X(\omega)$ *está en* B*.*

Definición 5. ***(Variable Continua)***
Sea X *una variable aleatoria, se dice que es una variable aleatoria continua cuando esta asociada a una función de densidad de probabilidad que a la vez está asociada a la función de probabilidad* P_X*.*

En general, se considera una variable aleatoria continua si toma valores a lo largo de un continuo, esto es, en todo un intervalo de valores.

[5] En general, una variable aleatoria toma más de un posible valor con una probabilidad de que pertenezca a un conjunto boreliano. Restrepo, Carlos Julio. Teoría de la probabilidad y aplicaciones, Variables Aleatorias, Popayán, editorial Unicauca, 2006. Pág. 62.

Definición 6. ***(Función de densidad para una variable continua)***
Sea X una variable aleatoria continua. La función de densidad de probabilidad de X, denotada por f_X se define para funciones de probabilidad de la forma $P_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$ tal que existe una función $f_X : \mathbb{R} \longrightarrow \mathbb{R}$ que satisface las siguientes condiciones.

$$f_X(x) \geq 0; \ \forall x \in \mathbb{R}$$

$$\int_{-\infty}^{+\infty} f_X(x) = 1 \tag{1.3}$$

Observación 1.2.1:

1. Si X es una variable aleatoria continua y B cualquier conjunto boreliano $B \in \mathcal{B}(\mathbb{R})$ la probabilidad de B, $(P_X(B))$ está definida por:

$$P_X(B) = \int_B f_X(x)dx \tag{1.4}$$

   Usualmente se consideran las siguientes notaciones:

$$\begin{aligned} P_X((a,b)) &= P(a < X < b) \\ P_X([a,b)) &= P(a \leq X < b) \\ P_X((a,b]) &= P(a < X \leq b) \\ P_X([a,b]) &= P(a \leq X \leq b) \end{aligned}$$

2. Por las expresiones (1.3) y (1.4) se tiene en particular que la probabilidad del conjunto boreliano $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ es igual a uno, es decir

$$P_X(\mathbb{R}) = \int_{\mathbb{R}} f_X(x)dx = \int_{-\infty}^{+\infty} f_X(x)dx = 1$$

3. Si X es una variable aleatoria continua, la probabilidad de que la variable X tome un valor perteneciente al intervalo (x,h) donde $h = x + \triangle x$ considerando $\triangle x$ un valor lo suficientemente pequeño, esto es aproximadamente $f_X(x)\triangle x$. Esto se asume de la siguiente manera:

$$P_X((x, x+\triangle x)) = P(x < X < x + \triangle x) \approx f_X(x)\triangle x \tag{1.5}$$

Definición 7. ***(Función masa de probabilidad)***
Sea P_X una función de probabilidad de la forma $P_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$. Considérese que existe una función $p_X : \mathbb{R} \longrightarrow \mathbb{R}$ definida para todos los números reales $x \in \mathbb{R}$, tal que $p_X(x) = 0$; $\forall x \in \mathbb{R}$, excepto para un conjunto $C \in \mathbb{R}$ finito o infinito numerable donde los valores de $p_X(x)$ son positivos, y la probabilidad de cualquier evento boreliano $B \in \mathcal{B}(\mathbb{R})$, $P_X(B)$ puede calcularse a través de la función p_X mediante la suma:

$$P_X(B) = \sum_{x \in B \cap C} p_X(x) \tag{1.6}$$

Cuando se toma en particular $B = \mathbb{R} \in \mathcal{B}(\mathbb{R})$, la expresión (1.6) tiene sentido, al imponer la condición de normalización.

$$P_X(\mathbb{R}) = \sum_{x \in \mathbb{R} \cap C} p_X(x) = \sum_{x \in C} p_X(x) = \sum_{x \in C} P(X = x) = 1 \tag{1.7}$$

Se dice que la función p_X es la función de masa de probabilidad de la variable X o de la función de probabilidad P_X, cuando la función de probabilidad P_X puede representarse en la forma (1.6).

Definición 8. ***(Variable discreta)***
Sea X una variable aleatoria, se dice que X es una variable aleatoria discreta cuando tiene asociada una función de masa de probabilidad p_X que satisface las expresiones (1.6) y (1.7).

Definición 9. ***(Función de distribución de probabilidad)***
Sea X una variable aleatoria. Considérese una función $F_X : \mathbb{R} \longrightarrow [0,1]$ definida para $\forall x \in \mathbb{R}$ de la siguiente manera:

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]) \tag{1.8}$$

A está función F_X se le llama función de distribución de probabilidad o distribución acumulativa asociada a la variable aleatoria X.

- Sea X una variable aleatoria discreta, la función de distribución F_X asociada a la variable X se determina a partir de la expresión (1.6) y está dada por

$$F_X(x) = P(X \leq x) = \sum_{u \in (-\infty, x] \cap C} p_X(u); \ \forall x \in \mathbb{R} \tag{1.9}$$

- Sea X una variable aleatoria continua, la función de distribución de probabilidad F_X asociada a la variable X se obtiene a partir de la expresión (1.4) y está dada por

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du; \ \forall x \in \mathbb{R} \tag{1.10}$$

Observación 1.2.2:

1. Sea X una variable aleatoria continua, f_X su correspondiente función de densidad y F_X su función de distribución de probabilidad. La función de densidad f_X puede obtenerse a partir de F_X, ya que la gráfica de $y = F_X(x)$ es una curva sin cortes, F_X es creciente y continua, y además F_X es diferenciable en todos los puntos (excepto, quizás en un número finito de ellos) y se obtiene de la siguiente manera:

$$F'_X(x) = \frac{d}{dx} F_X(x) = f_X(x) \tag{1.11}$$

2. Sea X una variable aleatoria discreta especificada por una función de masa de probabilidad p_X, F_X su correspondiente función de distribución de probabilidad dada en la expresión (1.9). La gráfica de $y = F_X(x)$, consiste de una sucesión de segmentos rectos horizontales, cada uno de los cuales es más alto que su predecesor, es decir, la gráfica de una función escalonada. La función F_X puede escribirse como

$$F_X(x) = \sum_{x_i \in C} p_X(x_i) u(x - x_i); \ \forall x \in \mathbb{R} \tag{1.12}$$

donde $u(x)$ es la función escalón unitario u definida:

$$u(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1, & \text{si } x \geq 0 \end{cases}$$

La función F_X puede expresarse de esta forma si hay presencia de discontinuidad.

3. Las propiedades que debe cumplir la función de distribución de probabilidad F_X para la variable aleatoria X:

   a) $F_X : \mathbb{R} \longrightarrow [0, 1]$ es una función creciente.

      Sean x_1 y x_2 dos números cualesquiera de $\mathbb{R}$ tal que $x_1 \leq x_2$, se tiene: $F_X(x_1) \leq F_X(x_2)$.

   b) Los limites de F_X, cuando x tiende a $+\infty$ ó $-\infty$, deben existir y ser

$$\lim_{x \to -\infty} F_X(x) = 0 \ ; \ \lim_{x \to +\infty} F_X(x) = 1$$

.

   c) En cualquier punto $x \in \mathbb{R}$, el limite cuando b toma valores positivos, existe

$$\lim_{b \to x^+} F_X(b) = F_X(x)$$

.

      Esto quiere decir, que F_X es continua por la derecha en todo punto $b \in \mathbb{R}$.

   d) En cualquier punto $x \in \mathbb{R}$, el limite cuando a toma valores negativos, debe ser

$$\lim_{a \to x^-} F_X(a) = F_X(x) - p_X(x)$$

      donde se define $p_X(x)$ como la probabilidad de la variable aleatoria X evaluada en el valor x. Se tiene $P_X(x) = P(X = x) = p_X(x)$. El valor $p_X(x)$ representa el tamaño del salto de la gráfica de $y = F_X(x)$ en x.

   e) Consideremos la función de distribución $F_X : \mathbb{R} \longrightarrow [0, 1]$, creciente y continua por la derecha; y se define una medida de probabilidad llamada Medida de Lebesgue - Stieltjes tal que $P_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0, 1]$ dada por:

$$P_X((a, b]) = F_X(b) - F_X(a) \tag{1.13}$$

      En la expresión (1.13) se establece una relación biyectiva entre las medidas Lebesgue y las funciones de distribución. De aqui se concluye que para especificar la función de probabilidad P_X basta con especificar la función de distribución de probabilidad F_X.

4. Sean X una variable aleatoria, F_X una función de distribución y p_x una función de masa asociadas a la variable X. Sean $a, b \in \mathbb{R}$ dos números cualquiera tal que $a < b$. Las probabilidades de los conjuntos borelianos

$(a,b),(a,b],[a,b),[a,b] \in B(\mathbb{R})$ en términos de las funciones F_X y p_X están dadas por las expresiones:

$$\begin{aligned}
P_X((a,b]) &= P(a < X \leq b) = F_X(b) - F_X(a) \\
P_X([a,b]) &= P(a \leq X \leq b) = F_X(b) - F_X(a) + p_X(a) \\
P_X([a,b)) &= P(a \leq X < b) = F_X(b) - F_X(a) + p_X(a) - p_X(b) \\
P_X((a,b)) &= P(a < X < b) = F_X(b) - F_X(a) - p_X(b)
\end{aligned}$$

Definición 10. ***(Función de probabilidad condicional)***
Sean X e Y dos variables aleatorias continuas definidas en un mismo espacio de probabilidad. Si $F(x,y)$ es la función de probabilidad conjunta tal que $F(x,y) = P(X \leq x,\ Y \leq y)$, derivable con respecto a sus argumentos, $f(x,y)$ la función de densidad de probabilidad conjunta entonces las funciones de probabilidades condicionales estan dadas de la siguiente manera:

- *La función de probabilidad condicional de X dado $Y = y$ es:*

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

- *La función de probabilidad condicional de Y dado $X = x$ es:*

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

donde,

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{+\infty} f(x,y)dx \\
f_X(x) &= \int_{-\infty}^{+\infty} f(x,y)dy
\end{aligned}$$

Definición 11. ***(Esperanza Matemática o Valor Esperado)***
Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y X una variable aleatoria. La esperanza matemática o valor esperado o valor medio de la variable aleatoria X, se denota $E[X] = \mu$, y se define:

- *Si X es una variable aleatoria discreta que puede tomar los diferentes valores $x_1, \dots, x_k$ con función de masa de probabilidad p_X tal que $p_X(x_i) = P(X = x_i)$, entonces la esperanza matemática de X se define:*

$$E[X] = \sum_{i=1}^{k} x_i p_X(x_i) = \sum_{i=1}^{k} x_i P(X = x_i); \quad \forall x \in \mathbb{R} \tag{1.14}$$

- *Si X es una variable aleatoria continua en cualquier subconjunto B, $B \in \mathcal{B}(\mathbb{R})$, con función de densidad f_X, entonces la esperanza matemática de X se define:*

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x); \quad \forall x \in \mathbb{R} \tag{1.15}$$

Entre las propiedades a considerar de la esperanza matemática, se tiene:

a) Si $(X \geq 0)$, entonces $E[X] \geq 0$.

b) Si $E[a + bX]$ existe con a y b constantes, entonces $E[a + bX] = a + bE[X]$.

c) $|E[X]| = E[|X|]$.

Definición 12. ***(Probabilidad Condicional)***
Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad (Ω un conjunto finito), $A \in \mathcal{A}$ cualquier subconjunto. La probabilidad condicional del evento B ($B \in \mathcal{A}$) asumiento que el evento A ha ocurrido y que tiene probabilidad $P(A) > 0$, denotada por $P(B|A)$, se define

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{1.16}$$

Algunas propiedades de la probabilidad condicional:

- $P(A|A) = 1$
- $P(\emptyset|A) = 0$
- $P(B|A) = 1$; cuando $A \subset B$.

Existen modelos de distribuciones de probabilidad para variables discretas y continuas. Examinaremos el concepto de la distribución Bernoulli para variables discretas, y la distribución de probabilidad para variables continuas como la distribución uniforme, y la normal. Aunque, en el documento no se emplea la distribución normal, se presenta con la finalidad de que el lector pueda identificar las posibles diferencias que existen entre los métodos de análisis paramétricos y no-paramétricos.

Definición 13. ***(La Distribución Bernoulli)***
Es una distribución de probabilidad discreta que toma valor 1 *para la probabilidad exito* (p) *y valor* 0 *para la probabilidad de fracaso* $(q = 1 - p)$.

Si X *es una variable aleatoria discreta que mide el número de éxitos, y se realiza un experimento con dos posibles resultados (exito o fracaso), se dice que la variable aleatoria* X *se distribuye como Bernoulli de parámetro* p, $X'Be(p)$,*tal que se tiene la siguiente formula.*

$$f_X(x) = p^x(1-p)^{1-x} \qquad con \quad x = \{0, 1\}$$

Definición 14. ***(La Distribución Uniforme)***
Se dice que una variable aleatoria X *continua tiene una distribución uniforme en el subconjunto* $[a, b] \in B(\mathbb{R})$, *si su función de densidad de probabilidad está dada por*

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & si\ a \leq x \leq b \\ 0, & para\ el\ resto \end{cases} \tag{1.17}$$

La función de distribución, se define:

$$F_X(x) = \begin{cases} 0, & si\ x < a \\ \frac{x-a}{b-a}, & para\ a \leq x \leq b \\ 1, & si\ x > b \end{cases} \tag{1.18}$$

Nota: *Si* X *es una variable aleatoria uniforme en cualquier subconjunto* $B \in B(\mathbb{R})$, *en particular* $[a, b]$, *se tiene:*

$$E[X] = \frac{a+b}{2}$$

Definición 15. ***(La Distribución Normal)***
Se dice que una variable aleatoria X sigue una distribución normal de parámetros μ y σ^2, lo que representamos del modo $X \sim N(\mu, \sigma^2)$, $|\mu| < \infty$, $\sigma^2 > 0$, si su función de densidad está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2\sigma^2}(x-\mu)^2};\ \forall x \in \mathbb{R} \tag{1.19}$$

La forma de la gráfica de la función de densidad es la llamada Campana de Gauss, que depende de los parámetros μ y σ^2. Además, El área contenida entre la gráfica y el eje de abscisas vale 1.

Observación 1.2.3

1. Consideremos $\mu = 0$ y $\sigma^2 = 1$, se dice que es una distribución normal tipificada o estandar $Z \sim N(0, 1)$, donde

   - La función de densidad está dada por:

   $$f_Z(z) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2\sigma^2}(z)^2};\ \forall z \in \mathbb{R}$$

   - La función de distribución está dada por:

   $$\Phi(z) = P(X \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2\sigma^2}(u)^2 du};\ \forall z \in \mathbb{R}$$

   Nota: *el simbolo ϕ representa la funcion de distribución de probabilidad de la distribución normal estandar.*

2. En el caso de que tengamos una distribución normal $X \sim N(\mu, \sigma^2)$ es posible realizar un cambio para obtener una distribución estandar $Z \sim N(0, 1)$. El cambio se efectua de la siguiente manera,

   $$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \rightsquigarrow N(0, 1)$$

3. Consideremos $X \sim N(\mu, \sigma^2)$. Para calcular la función de distribución $F_X(x) = P(X \leq x)$, es necesario tener presente que dados $a, b \in \mathbb{R}$, se tiene que

   $$X \sim N(\mu, \sigma^2) \Rightarrow Y = a + b$$

Luego,

$$X \sim (a + b\mu, (b\sigma)^2)$$

De aquí, hacemos el cambio

$$Z = \frac{X - \mu}{\sigma} N(0, 1)$$

y calculamos $z = \frac{x-\mu}{\sigma}$. Recurriendo a la tabla de distribución normal estandar $N(0, 1)$, se tiene $F_Z(z) = P(Z \leq z)$.

Pero, como

$$P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(X \leq x) = F_X(x)$$

1.3. Función de supervivencia.

En esta sección definiremos una función que aporta información sobre la probabilidad de supervivencia de las unidades de una muestra tomada a una población en estudio.

Consideremos Ω el espacio muestral conformado por los tiempos de supervivencia de las unidades de la muestra en estudio, considerando que las unidades ingresaron al estudio desde un instante de tiempo inicial $x = 0$, y $X : \Omega \longrightarrow \mathbb{R}^+$ la variable aleatoria continua que representa *el tiempo de supervivencia de una unidad.*

Como X es una variable aleatoria continua que representa el tiempo de supervivencia de una unidad y F_X su correspondiente función de distribución de probabilidad que está dada de la siguiente manera:

$$F_X(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ P(X \leq x), & \text{si } x > 0 \end{cases}$$

donde, $F_X(x) = P(X \leq x)$, es la probabilidad que tiene una unidad que esta siendo observada, de experimentar el evento de interés antes de alcanzar el instante de tiempo x. Consideremos la función de probabilidad S_X definida así:

$$\begin{aligned} S_X : \mathbb{R}^+ \cup \{0\} &\longrightarrow [0, 1] \\ x &\longrightarrow S_X(x) = P(X > x) = 1 - F_X(x) \end{aligned} \tag{1.20}$$

donde, $S_X(x) = P(X > x)$ es la probabilidad que tiene una unidad que esta siendo observada, de experimentar el evento de interés después del instante de tiempo x; esta función S_X es llamada **Función de Supervivencia**.

La función S_X satisface las siguientes propiedades:

1. $S_X(0) = 1$.

   Demostración. De la expresión (1.23), para $x = 0$,

   $$S_X(0) = 1 - F_X(0)$$

   por definición de la función de distribución,

   $$F_X(0) = 0$$

   Por consiguiente,

   $$S_X(0) = 1$$

   □

   Esto indica que ninguna unidad de la muestra en estudio experimentó el evento de interés en el inicio del estudio, es decir, en el instante de tiempo $x = 0$.

2. S_X es una función decreciente.

   Demostración. Como F_X es una función creciente, por la expresión (1.23), se tiene que S_X es decreciente.

   Esto significa que a medida que una unidad en estudio incrementa su tiempo de supervivencia, la probabilidad de no experimentar el evento de interés del estudio disminuye. □

3. Si existe el instante de límite ω en el estudio, $(\omega \in \mathbb{R}^+)$, entonces $S_X(\omega) = 0$.

   Supongamos que existe por lo menos un $\omega \in \mathbb{R}^+$ tal que,
   $\omega = \inf\{x > 0/P(X > x) = 0\}$ entonces $S_X(\omega) = 0$.

Demostración. Consideremos $\omega \in \mathbb{R}^+$. Por la expresión (1.23)

$$S_X(\omega) = 1 - F_X(\omega)$$

Como $\omega = \inf\{x > 0/P(X > x) = 0\}$

$$\begin{aligned} S_X(\omega) &= \lim_{x \to \omega^+} S_X(x) \\ &= \lim_{x \to \omega^+} [1 - F_X(x)] \\ &= 1 - \lim_{x \to \omega^+} F_X(x) \end{aligned}$$

Como F_X es una función creciente en $[0, 1]$ y es continua por la derecha

$$\lim_{x \to \omega^+} F_X(x) = F_X(\omega)$$

Por consiguiente,

$$S_X(\omega) = 1 - F_X(\omega)$$

Si ω es el instante de tiempo límite en el cual no se puede experimentar el evento de interés, $F_X(\omega) = 1$

Por lo tanto,

$$S_X(\omega) = 0$$

Esto quiere decir, que en el instante de tiempo ω, el período de observación o tiempo de seguimiento ya culminó, de hecho, se espera que las unidades objeto de estudio hayan experimentado el evento de interés antes del instante de tiempo ω. $\square$

La gráfica de la función de supervivencia S_X en función del tiempo recibe el nombre de **Curva de Supervivencia**. Por ejemplo, en la gráfica 1.1, el par de valores $t = 90$, $S_X(90) = 0{,}30$ indica que la probabilidad de que una unidad del estudio experimente el evento de interés más alla del instante de tiempo 90, es aproximadamente del 30 %.

GRÁFICA 1.1 Curva de Supervivencia.

Fuente: *PREDOMINGO, Alejandro. Fundamentos Teóricos del Análisis de Sobrevivencia. Pág. 491.*

Observación 1.3.1
Sea X una variable aleatoria continua, si F_X es diferenciable, entonces por la expresión (1.11) se tiene que la función de densidad de probabilidad de la variable X, denotada f_X, está dada de la siguiente manera:

$$f_X(x) = \begin{cases} 0, & \text{si } x \leq 0 \\ F'_X(x), & \text{si } x > 0 \end{cases}$$

donde, $F'_X(x) = \frac{d}{dx} F_X(x)$. Por la expresión (1.23),

$$f_X(x) = -S'_X(x) \quad \text{para } x > 0$$

donde, $S'_X(x) = \frac{d}{dx} S_X(x)$.

1.4. La fuerza de mortalidad.

Sea X una variable aleatoria continua no negativa ($X \geq 0$), y supongamos que F_X es diferenciable. La fuerza de mortalidad en el instante de tiempo x, se denota

$\mu_X(x)$, y es una medida de la fuerza o intensidad de la mortalidad en el instante de tiempo x, que se define de la siguiente manera:

$$\mu_X(x) = \frac{f_X(x)}{S_X(x)} \qquad \text{para} \quad 0 \leq x < \omega \tag{1.21}$$

Como F_X es diferenciable, y $f_X(x) = F'_X(x)$ para $x > 0$

$$\mu_X(x) = \frac{F'_X(x)}{S_X(x)}$$

De la expresión (1.20),

$$\mu_X(x) = \frac{-\frac{d}{dx}(S_X(x))}{S_X(x)}$$

Por consiguiente,

$$\mu_X(x) = -\frac{S'_X(x)}{S_X(x)} \quad \text{para} \quad 0 \leq x < \omega \tag{1.22}$$

donde, $S'_X(x) = \frac{d}{dx}(S_X(x))$.

$\mu_X(x)$ es la probabilidad de riesgo de que una unidad del estudio que no ha experimentado el evento de interés hasta el instante de tiempo x, lo experimente en un instante de tiempo $x + \triangle x$, considerando $\triangle x$ suficientemente pequeño. De la expresión (1.5) se tiene,

$$\mu_X(x) = P(x < X < x + \triangle x) \approx f_X(x)\triangle x$$

Ejemplo 1.1

Para la formulación del siguiente ejemplo, se tuvo en cuenta el ejemplo 2.1 presentado por Huertas, Jaime Abel en su libro Cálculo Actuarial: Contingencias de vida individual, pág. 38.

Un equipo de investigadores médicos están interesados en determinar el tiempo de supervivencia de un grupo de individuos después de haberles diagnosticado una cierta enfermedad.

Supongamos que la variable aleatoria X representa el tiempo que transcurre desde que se le diagnosticó al paciente cierta enfermedad hasta que le ocurre el deceso (muerte), la cual sigue una distribución uniforme en el intervalo de tiempo $[0, b]$, $b > 0$.

Solución:

Como X es una variable aleatoria que sigue una distribución uniforme en el intervalo $[0, b]$, $b > 0$, de la expresión (1.17):

$$f_X(x) = \begin{cases} \frac{1}{b-0}, & \text{si } 0 \leq x \leq b \\ 0, & \text{en cualquier otro caso} \end{cases}$$

Por la expresión (1.10),

$$F_X(x) = \int_0^x f_X(x)dx = \int_0^x \frac{1}{b}dx = \frac{x}{b} \qquad \text{para} \quad 0 \leq x \leq b$$

Luego,

$$F_X(x) = \begin{cases} 0, & \text{si } x < 0 \\ \frac{x}{b}, & \text{si } 0 \leq x \leq b \\ 1, & \text{si } x > b \end{cases}$$

Por lo tanto, de la expresión (1.20)

$$S_X(x) = 1 - \frac{x}{b} \qquad \text{para} \quad x \in [0, b]$$

Verifiquemos si la función S_X cumple las propiedades de una función de supervivencia:

- $S_X(0) = 1 - \frac{0}{b} = 1 - 0 = 1$
- Teniendo en cuenta que $S'_X(x) = -\frac{1}{b} < 0,\ b > 0$ entonces S_X es decreciente[6] en el intervalo $[0, b]$.
- $S_X(b) = 1 - \frac{b}{b} = 1 - 1 = 0$

[6]Por el criterio de la primera derivada.

Por consiguiente, la función $S_X(x) = 1 - \frac{x}{b}$, para $x \in [0, b]$ cumple con las propiedades de una función de supervivencia.

Con base a las funciones obtenidas de supervivencia, función de densidad, función de distribución con respecto a la variable aleatoria X, calculemos las siguientes probabilidades:

1. La probabilidad de riesgo de que los individuos que están siendo observados desde que se les diagnosticó cierta enfermedad, les ocurra el evento de interés (la muerte), después de 12 años de tratamiento.

   De la expresión (1.21),

   $$\mu_X(x) = \frac{f_X(x)}{S_X(x)} = \frac{\frac{1}{b}}{1 - \frac{x}{b}} = \frac{1}{b - x} \quad para \quad 0 < x < b$$

   Supongamos que el equipo de investigadores se basan en otras investigaciones donde el tiempo máximo de supervivencia de los individuos que padecen la enfermedad objeto de estudio, y que siguen un tratamiento riguroso, es de aproximadamente 15 años. La variable aleatoria X sigue una distribución uniforme en el intervalo de tiempo $[0, 15]$. De aqui, se obtiene

   $$\mu_X(12) = \frac{1}{15 - 12} = 0{,}33$$

   Este valor indica, que en el 33,33 % de los casos se espera que los individuos que están siendo observados desde el momento que se les diagnosticó cierta enfermedad, les ocurra el evento de interés, la muerte, después de 12 años de tratamiento, dado que se espera que no experimenten la muerte antes de 15 años de tratamiento.

2. La probabilidad que tiene un individuo del grupo en estudio de experimentar el evento de interés, la muerte, después del instante de tiempo t y antes del instante de tiempo k, si $0 < t < k$.

   $$P(t < X \leq k) = \int_t^k f_X(x)dx = \int_t^k \frac{1}{b}\; dx = \frac{x}{b}\Big|_t^k = \frac{k - t}{b}$$

De igual forma, puede obtenerse a partir de la expresión (1.13)

$$P(t < X \leq k) = P(X > t) - P(X > k) = F_X(k) - F_X(t)$$

Como $F_X(x) = \frac{x}{b}$ para $0 \leq x < b$

$$P(t < X \leq k) = F_X(k) - F_X(t) = \frac{k}{b} - \frac{t}{b} = \frac{k-t}{b}$$

3. La probabilidad de que un individuo del grupo en estudio alcance el instante de tiempo k, dado que cuando ingreso al período de observación ya había alcanzado el instante de tiempo t, si $(0 < t < k)$.

   Si $k > t$, se debe calcular la siguiente probabilidad: $P(X > k|X > t)$

   Por la expresión (1.16),

$$P(X > k|X > t) = \frac{P[(X > k) \cap (X > t)]}{P(X > t)}$$

   Como $k > t,\ (X > k) \cap (X > t) = (X > k)$

   Luego,

$$\begin{aligned} P(X > k|X > t) &= \frac{P(X > k)}{P(X > t)} \\ &= \frac{S_X(k)}{S_X(t)} \\ &= \frac{1 - \frac{k}{b}}{1 - \frac{t}{b}} \\ &= \frac{b-k}{b-t}. \end{aligned}$$

   Para interpretar este resultado, consideremos que el tiempo máximo de supervivencia de los individuos que siguen un tratamiento riguroso después del diagnóstico es de aproximadamente 15 años. Sea $t = 12$ años, la edad en que se le diagnosticó la enfermedad a un individuo, la probabilidad de que el individuo no experimente la muerte antes de alcanzar el instante de tiempo de los 20 años, es del aproximadamente 46,67 % cuando se espera que no experimente la muerte hasta los 27 años.

$$P(X > 20|X > 12) = \frac{27-20}{27-12} = \frac{7}{15} \approx 46{,}67$$

1.5. Probabilidades Condicionales de Supervivencia.

De un conjunto de tiempos de supervivencia x, $(x > 0)$ (incluyendo los tiempos censurados) de una muestra de unidades tomada de la población en estudio, se pueden obtener a partir de las funciones de supervivencia, de densidad, y fuerza de mortalidad de la variable aleatoria X, la proporción de unidades en un estudio que experimentarán el evento de interés durante un intervalo de tiempo.

Sea $x > 0$, fijo pero arbitrario y Θ el espacio muestral conformado por los tiempos de supervivencia de un grupo de unidades que han alcanzado el instante de tiempo x, y $T_x : \Theta \longrightarrow \mathbb{R}^+$ la variable aleatoria continua que representa *el tiempo de supervivencia de una unidad que ha alcanzado el instante de tiempo* x, esto se expresa de la siguiente manera

$$T_x = X - x$$

donde, T_x se interpreta como el lapso de tiempo que le queda a la unidad en estudio, para experimentar el evento de interés o cualquier otro, dado que ha alcanzado el instante de tiempo x.

Si se considera que existe el instante de tiempo límite ω para experimentar el evento de interés, la variable T_x así definida, toma valores dentro del intervalo $(0,\ \omega - x)$. De lo contrario, si no existe un instante de tiempo límite, T_x toma valores en intervalo $(0, \infty)$.

Teniendo en cuenta que T_x es una variable aleatoria continua, la de $X - x$, condicionada por el evento $(X > x)$, esto es, por $T_x > 0$, la función de distribución de probabilidad de la variable T_x, denotadá F_{T_x}; para $x > 0$ fijo, y $t > 0$, está dada mediante la siguiente expresión:

$$F_{T_x}(t) = \begin{cases} 0, & \text{si } t \leq 0 \\ P(T_x \leq t), & \text{si } t > 0 \end{cases}$$

Si $F_{T_x}(t) = P(T_x \leq t)$ para $t > 0$, F_{T_x} puede expresarse de la siguiente manera:

$$
\begin{aligned}
F_{T_x}(t) = P(T_x \leq t) &= P(X - x \leq t | X > x) \\
&= P(X \leq x + t | X > x) \\
&= \frac{P((X \leq x + t) \cap (X > x))}{P(X > x)}, \qquad \text{por (1.16)} \\
&= \frac{P(x < X \leq x + t)}{P(X > x)}, \qquad 0 < x < t, \quad x < x + t \\
&= \frac{P(X > x) - P(X > x + t)}{P(X > x)}, \qquad \text{por (1.13)} \\
&= \frac{S_X(x) - S_X(x + t)}{S_X(x)}, \qquad \text{por (1.20)} \\
&= 1 - \frac{S_X(x + t)}{S_X(x)}
\end{aligned}
$$

Luego,

$$F_{T_x}(t) = 1 - \frac{S_X(x + t)}{S_X(x)}$$

puede expresarse por medio de la notación actuarial;

$$F_{T_x}(t) = 1 - \frac{S_X(x + t)}{S_X(x)} = {}_tq_x \tag{1.23}$$

La notación ${}_tq_x$, se explica por sí misma [7]. El postsubíndice, x, indica el instante de tiempo que ha alcanzado la unidad, y el presubíndice, t, la longitud del intervalo temporal para el que se calcula la probabilidad del evento de interés.

donde ${}_tq_x$, es la probabilidad de que una unidad del grupo en estudio, experimente el evento de interés entre los instantes de tiempos x y $x + t$, dado que la unidad ha alcanzado el instante de tiempo x. Es decir,

$${}_tq_x = P(X \leq x + t | X > x)$$

A partir de la expresión (1.23), se puede demostrar que ${}_tq_x = F_{T_x}(t)$.

[7] Parte de la notación que se utiliza actualmente fue adoptada tras el congreso Internacional de Actuarios que tuvo lugar en 1,898.

Como $F_{T_x}(t) = P(T_x \leq t)$,

$$\begin{aligned} F_{T_x}(t) &= P(X - x \leq t), \quad \text{si} \quad T_x = X - x \\ &= P(X \leq x + t) \\ &= F_X(x + t), \quad \text{por definición de } F_X. \end{aligned}$$

Luego,

$$F_{T_x}(t) = F_X(x + t) \tag{1.24}$$

En está última expresión, vale decir, que no importa la condición del evento $(X > x)$, dado que

$$F_X(x + t) = 1 - S_X(x + t)$$

Como X esta definida para $\forall x \in [0, +\infty)$, si consideramos, $x = 0$ y $t > 0$,

$$F_X(0 + t) = 1 - S_X(0 + t)$$

Esto indica, que la función de distribución de probabilidad de la variable X, evaluada en cualquier instante de tiempo $x + t$, con $x \geq 0$ y $t > 0$, toma valores en el intervalo $(0, +\infty)$, que es lo mismo que considerar la función de distribución de probabilidad de la variable T_x; esto es,

$$F_X(x + t) = F_{T_x}(t)$$

Ahora, si evaluemos la función de supervivencia S_X, en el instante de tiempo $x + t$, $S_X(x + t)$.

$$\begin{aligned} S_X(x + t) &= 1 - F_X(x + t), \qquad \text{por} \quad (1{,}20) \\ &= 1 - F_{T_x}(t), \qquad \text{por} \quad (1{,}24) \\ &= S_{T_x}(t) \end{aligned}$$

Luego,

$$S_X(x + t) = S_{T_x}(t) \tag{1.25}$$

De la expresión (1.25), se tiene;

$$\begin{aligned} S_X(x) &= S_X(x+0) \\ &= S_{T_x}(0) \\ &= 1 - F_{T_x}(0) \end{aligned}$$

Como $F_{T_x}(0) = 0$

$$S_X(x) = 1$$

Por consiguiente,

$$\begin{aligned} {}_tq_x &= 1 - \frac{S_X(x+t)}{S_X(x)} \\ &= 1 - \frac{S_X(x+t)}{1} \\ &= 1 - S_X(x+t) \end{aligned}$$

por la expresión (1.25)

$${}_tq_x = 1 - S_{T_x}(t)$$

Por lo tanto,

$${}_tq_x = F_{T_x}(t)$$

A través de la demostración anterior, puede notarse que;

$$\begin{aligned} 1 - F_{T_x}(t) &= 1 - P(X \leq x+t); \qquad \text{si} \quad T_x = X - x. \\ &= P(X > x+t); \quad \text{por def. de complemento.} \\ &= S_X(x+t) \\ &= S_{T_x}(t) \end{aligned}$$

Luego, $1 - F_{T_x}(t) = S_{T_x}(t)$ es la probabilidad del suceso contrario al que define la probabilidad $F_{T_x}(t) = {}_tq_x$, por lo que valdrá,

$$S_{T_x}(t) = 1 - {}_tq_x$$

donde, $S_{T_x}(t) = {}_tp_x$ por notación actuarial. En la notación ${}_tp_x$, El postsubíndice, x, indica el instante de tiempo que ha alcanzado la unidad, y el presubíndice, t, la longitud del intervalo temporal para el que se calcula la probabilidad del evento de interés.

${}_tp_x$, es la probabilidad de que una unidad del grupo en estudio, no experimente el evento de interés antes del instante de tiempo $x + t$, dado que ha alcanzado el instante de tiempo x. Es decir,

$$ {}_tp_x = P(X > x + t | X > x) $$

Como ${}_tp_x = 1 - {}_tq_x$, se tiene de la expresión (1.23)

$$ \begin{aligned} {}_tp_x &= 1 - \left(1 - \frac{S_X(x+t)}{S_X(x)}\right) \\ &= \frac{S_X(x+t)}{S_X(x)} \end{aligned} $$

Luego,

$$ {}_tp_x = \frac{S_X(x+t)}{S_X(x)} \tag{1.26} $$

Nota: *Las expresiones, ${}_tq_x$ y ${}_tp_x$ pueden interpretarse de la siguiente manera cuando se calculan para un grupo en general: ${}_tq_x$ como la proporción de unidades que experimentaron el evento de interés del estudio durante el intervalo de tiempo $(x,\ x+t)$, dado que las unidades habían alcanzado el instante de tiempo x, y ${}_tp_x$ como la proporción de unidades que no experimentaron el evento de interés hasta el instante de tiempo $x+t$, dado que habían alcanzado el instante de tiempo x.*

Observación 1.5.1

1. Sea T_x una variable aleatoria continua no negativa ($T_x \geq 0$), para todo $x > 0$ fijo, y su correspondiente función de distribución de probabilidad F_{T_x} es diferenciable, la función de densidad de probabilidad de la variable T_x, denotada f_{T_x}, se deduce de la siguiente manera:

Como $F_{T_x}(t) = P(T_x \leq t) = {}_tq_x, \quad$ para $t > 0,$ se sigue que;

$$\begin{aligned}
f_{T_x}(t) &= \frac{d}{dt}({}_tq_x) \quad \text{por la expresión (1.11)} \\
&= \frac{d}{dt}(1 - {}_tp_x) \\
&= \frac{d}{dt}\left(1 - \frac{S_X(x+t)}{S_X(x)}\right) \\
&= \frac{d}{dt}(1 - S_X(x+t)); \qquad S_X(x) = 1 \\
&= -\frac{d}{dt}S_X(x+t)
\end{aligned}$$

De la expresión (1.25),

$$f_{T_x}(t) = -S'_{T_x}(t) \tag{1.27}$$

donde, $-\frac{d}{dt}S_{T_x}(t) = -S'_{T_x}(t)$.

A partir de la expresión (1.27), se obtiene una expresión de la función de densidad de probabilidad de la variable T_x en términos de la probabilidad ${}_tp_x$ y de la fuerza de mortalidad de la variable X, en el instante de tiempo $x+t$.

$$\begin{aligned}
f_{T_x}(t) &= -S'_{T_x}(t) \\
&= -S'_X(x+t); \qquad \text{por} \quad (1{,}25)
\end{aligned}$$

Como $S_X(x) = 1$

$$f_{T_x}(t) = -\frac{S'_X(x+t)}{S_X(x)}$$

Si en esta última expresión, a la derecha múltiplicamos y dividimos por $S_X(x+t)$, cuando $x+t > 0$, se tiene:

$$\begin{aligned}
f_{T_x}(t) &= \frac{-S'_X(x+t)}{S_X(x)}\left(\frac{S_X(x+t)}{S_X(x+t)}\right) \\
&= \frac{S_X(x+t)}{S_X(x)}\left(\frac{-S'_X(x+t)}{S_X(x+t)}\right)
\end{aligned}$$

Por las expresiones (1.26) y (1.22),

$$f_{T_x}(t) = {}_tp_x \ \mu_X(x+t) \qquad \text{para} \quad x+t>0 \tag{1.28}$$

2. Mediante la definición de la función de densidad de probabilidad de la variable T_x, para $x>0$ fijo, pueden estimarse algunas probabilidades, de la siguiente manera:

   a) $P(x<X<x+t) = \int_0^t f_{T_x}(t)\ dt$

   $$\begin{aligned}\int_0^t f_{T_x}(t)\ dt &= -\int_0^t S'_X(x+t)\ dt\\ &= -S_X(x+t)|_0^t\\ &= -(S_X(x+t)-S_X(x+0))\\ &= -S_X(x+t)+S_X(x)\\ &= 1-S_X(x+t)\\ &= 1-S_{T_x}(t)\\ &= F_{T_x}(t)\end{aligned}$$

   De la expresión (1.23),

   $$F_{T_x}(t) = {}_tq_x$$

   b) Para cada $u>0$, la probabilidad de que una unidad del estudio experimente el evento de interés durante el intervalo de tiempo $(x+t,\ x+t+u)$, dado que ha alcanzado el instante de tiempo x, se denota ${}_{t/u}q_x$. Esto es,

   $${}_{t/u}q_x = P(t<T_x<t+u) = \frac{P(x+t<X\leq x+t+u)}{P(X>x)}$$

   $$\begin{aligned}P(t<T_x<t+u) &= \int_t^{t+u} f_{T_x}(y)\ dy; \quad \text{para} \quad y=t+u\\ &= -\int_t^{t+u} S'_X(x+y)\ dy\\ &= -S_X(x+y)|_t^{t+u}\\ &= -(S_X(x+y)-S_X(x+t))\\ &= -S_X(x+y)+S_X(x+t)\\ &= -S_{T_x}(y)+S_{T_x}(t) \qquad por \quad (1{,}25)\\ &= -{}_yp_x+{}_tp_x\end{aligned}$$

Como $\;{}_yp_x = 1 - {}_yq_x, \qquad {}_tp_x = 1 - {}_tq_x$

$$P(t < T_x < t+u) = -{}_tq_x + {}_yq_x$$

Por lo tanto,

$${}_{t/u}q_x = {}_{t+u}q_x - {}_tq_x$$

También, la expresión ${}_{t/u}q_x$, puede expresarse de las siguientes maneras:

- ${}_{t/u}q_x = {}_tp_x \cdot {}_uq_{x+t}$

  Está expresión, puede demostrarse así:

  Como,

  $${}_{t/u}q_x = \frac{P(x+t < X \leq x+t+u)}{P(X > x)}$$

  Considerando que $P(X > x+t) \neq 0$

  $$\begin{aligned}
  {}_{t/u}q_x &= \frac{P(x+t < X \leq x+t+u)}{P(X > x)} \cdot \frac{P(X > x+t)}{P(X > x+t)} \\
  &= \frac{P(X > x+t)}{P(X > x)} \cdot \frac{P(x+t < X \leq x+t+u)}{P(X > x+t)} \\
  &= \frac{S_X(x+t)}{S_X(x)} \cdot \frac{P(X > x+t) - P(X > x+t+u)}{P(X > x+t)} \qquad \text{por} \quad (1{,}13) \\
  &= \frac{S_X(x+t)}{S_X(x)} \cdot \left(1 - \frac{P(X > x+t+u)}{P(X > x+t)}\right) \\
  &= \frac{S_X(x+t)}{S_X(x)} \cdot \left(1 - \frac{S_X(x+t+u)}{S_X(x+t)}\right) \\
  &= {}_tp_x \cdot {}_uq_{x+t}
  \end{aligned}$$

- ${}_{t/u}q_x = {}_tp_x - {}_{t+u}p_x$

  Está expresión, puede demostrarse así:

  Como,

  $${}_{t/u}q_x = \frac{P(x+t < X \leq x+t+u)}{P(X > x)}$$

$$
\begin{aligned}
{}_{t/u}q_x &= \frac{P(X > x+t) - P(X > x+t+u)}{P(X > x)} \\
&= \frac{P(X > x+t)}{P(X > x)} - \frac{P(X > x+t+u)}{P(X > x)} \\
&= \frac{S_X(x+t)}{S_X(x)} - \frac{S_X(x+t+u)}{S_X(x)} \\
&= {}_tp_x - {}_{t+u}p_x
\end{aligned}
$$

donde, ${}_{t/u}q_x$ puede interpretarse, para que la unidad del estudio, experimente el evento de interés entre el intervalo de tiempo $(x+t,\ x+t+u)$, dado que ya alcanzo el instante de tiempo x, es necesario que alcance el instante de tiempo $x+t$, y experimente el evento de interés antes de u tiempo después, es decir, que experimente el evento dentro del intervalo $(x+t,\ x+t+u)$.

3. Nótese, que las probabilidades ${}_tq_x$ y ${}_tp_x$ están dadas por las funciones $F_{T_x}(t)$ y $S_{T_x}(t)$, las cuales están condicionadas por el evento $(X > x)$. Esto quiere decir,

$$
\begin{aligned}
F_{T_x}(t) &= F_X(x+t|X > x) \\
S_{T_x}(t) &= S_X(x+t|X > x)
\end{aligned}
$$

   a) Para la función de densidad de la variable T_x, $f_{T_x}(t)$, condicionada por el evento $(X > x)$, se tiene

$$
\begin{aligned}
f_{T_x}(t) &= f_X(y|X > x) \qquad \text{para} \quad y = x+t \\
&= \frac{d}{dy}F_X(y|X > x) \qquad \text{por definición de } f_X \\
&= \frac{d}{dy}\left(1 - \frac{S_X(y)}{S_X(x)}\right) \qquad \text{por} \quad (1,23) \\
&= \frac{d}{dy}\left(-\frac{S_X(y)}{S_X(x)}\right) \\
&= -\frac{S'_X(y)}{S_X(x)} \\
&= \frac{f_X(y)}{S_X(x)} \qquad \text{por definición de } f_X \\
&= \frac{f_X(x+t)}{S_X(x)}
\end{aligned}
$$

como $S_X(x) = 1$, se llega a la siguiente igualdad:

$$f_{T_x}(t) = f_X(x+t|X > x) = f_X(x+t)$$

En está última expresión, vale decir, que no importa la condición del evento $(X > x)$, dado que

$$f_X(x+t) = -S'_X(x+t)$$

Como X esta definida para $\forall x \in [0, +\infty)$, si consideramos, $x = 0$ y $t > 0$,

$$f_X(0+t) = -S'_X(0+t)$$

Esto indica, que la función de densidad de la variable X, evaluada en cualquier instante de tiempo $x+t$, con $(x \geq 0)$ y $(t > 0)$, toma valores en el intervalo $(0, +\infty)$, que es lo mismo que considerar la función de densidad de la variable T_x. Esto es,

$$f_X(x+t) = f_{T_x}(t)$$

Igualmente, para calcular la fuerza de mortalidad de la variable T_x, $\mu_{T_x}(t)$, basta con calcular $\mu_X(x+t)$.

b) La fuerza de mortalidad de la variable T_x, $\mu_{T_x}(t)$, condicionada por el evento $(X > x)$, se tiene

$$\begin{aligned} \mu_{T_x}(t) &= \mu_X(y|X > x) && \text{para} \quad y = x+t \\ &= \frac{f_X(y|X > x)}{S_X(y|X > x)} && \text{por} \quad (1{,}21) \\ &= \frac{\frac{f_X(y)}{S_X(x)}}{\frac{S_X(y)}{S_X(x)}} \\ &= \frac{f_X(y)}{S_X(y)} \\ &= \frac{f_X(x+t)}{S_X(x+t)} \end{aligned}$$

De la expresión (1.21), se tiene que:

$$\mu_{T_x}(t) = \mu_X(x+t|X > x) = \mu_X(x+t)$$

Ejemplo 1.2
El siguiente ejemplo fue tomado del profesor de la facultad de economia en la Universidad de Barcelona, Manuel Artis en su libro "Estadística Actuarial de Vida", pág. 208.

Considérese T_x la variable aleatoria que representa el tiempo de duración de un objeto automático. T_x tiene como función de distribución:

$$F_{T_x}(t) = \frac{t}{36-x}; \qquad 0 \leq t < 36-x, \qquad x \geq 36-x.$$

Verificar si se cumple: $1 - F_{T_x}(t) = S_{T_x}(t)$, donde $S_{T_x}(t)$ es una función de supervivencia.

***Solución**:*

Para verificar que S_{T_x} es una función de supervivencia, S_{T_x} debe satisfacer las siguientes propiedades:

1. $S_{T_x}(0) = 0$

   $S_{T_x}(0) = 1 - \frac{0}{36-x} = 0$

2. S_{T_x} sea una función decreciente.

   Como $S'_{T_x}(t) = -\frac{1}{36-x} < 0$, por el criterio de la primera derivada, se puede concluir que S_{T_x} es decreciente para $\forall\ t \in [0,\ 36-x)$.

3. Existe (o no) un instante de tiempo límite t, tal que $S_{T_x}(t) = 0$

   La función $S_{T_x}(t)$ esta definida para $t \in [0,\ 36-x)$. Cuando $t = 36-x$, $S_{T_x}(36-x) = 0$.

   Esto quiere decir, que en el instante de tiempo $t = 36-x$ se espera que el objeto falle sin la posibilidad de arreglo.

Por lo tanto,

$$S_{T_x} = 1 - \frac{t}{36-x}$$

es una función de supervivencia.

Para la variable T_x, la función de densidad de probabilidad, vale:

$$f_{T_x}(t) = F'_{T_x}(t) = \frac{d}{dt}\left(\frac{t}{36-x}\right) = \frac{1}{36-x}$$

Teniendo en cuenta la información obtenida, calculemos las siguientes probabilidades:

a. La probabilidad de que el objeto de estudio experimente algún daño en el sistema antes de transcurridos 20 meses de su elaboración, dado que se vende después de haber permanecido en el almacen $x = 10$ meses. Esto es ${}_{10}q_{10}$:

   De la expresión (1.23), se tiene:

$$ {}_{10}q_{10} = F_{T_{10}}(10) = \frac{10}{60-10}$$

   Otra forma de calcular ${}_{10}q_{10}$,

$$ {}_{10}q_{10} = \int_0^{10} f_{T_x}(t)dt.$$

b. La probabilidad de que el objeto de estudio, experimente algún daño en el sistema entre los 25 meses y los 35 meses, dado que el objeto se vende después de haber permanecido en el almacen $x = 20$ meses. Esto es, ${}_{5|10}q_{20}$:

   De la observación (1.5.1) numeral $2b$, se tiene:

$$ {}_{5|10}q_{20} = {}_{15}q_{10} - {}_{5}q_{10}$$

   Por la expresión (1.23),

$$ {}_{5|10}q_{20} = F_{T_{10}}(15) - F_{T_{10}}(5) = \frac{15}{36-10} - \frac{5}{36-10} = \frac{10}{36-10}$$

1.6. La Esperanza de vida

La esperanza de vida[8] o vida probable de las unidades de una muestra tomada a la población en estudio, es un valor que representa la cantidad de tiempo esperado de vida o tiempo promedio de vida para una unidad del grupo en estudio.

- La esperanza de vida o vida probable para una unidad del grupo en estudio, que ingresa al estudio desde el instante de tiempo inicial $x = 0$, se denota e_0^0, y está dada mediante la siguiente expresión:

$$e_0^0 = E[X] = \int_0^\infty x f_X(x) dx \tag{1.29}$$

La expresión (1.29), puede expresarse de la siguiente manera:

$$e_0^0 = E[X] = \int_0^\infty {}_x p_0 \; \mu_X(x) dx \tag{1.30}$$

Está expresión, puede demostrarse así:

$$\begin{aligned} e_0^0 = E[X] &= \int_0^\infty x f_X(x) dx \\ &= \int_0^\infty x(-S'_X(x)) dx \qquad \text{por la observación 1.3.1} \end{aligned}$$

Como $S_X(x) = 1$

$$e_0^0 = E[X] = \int_0^\infty x \left(\frac{-S'_X(x)}{S_X(x)} \right) S_X(x) dx$$

como $S_X(x)$, puede expresarse como $S_X(x) = 1 - F_{T_x}(0)$,

[8]El término vida, se utiliza para englobar las actividades características de todos los organismos (desde algas unicelulares hasta plantas y animales), espacio de tiempo que transcurre desde el nacimiento de una persona(animal o vegetal) hasta su muerte, duración de las cosas, acciones notables ejecutadas por una persona durante su vida. www.ladefinición.com

$$
\begin{aligned}
e_0^0 = E[X] &= \int_0^\infty x\mu_X(x)(1-F_{T_0}(x)); \qquad \text{por} \quad (1{,}22)\\
&= \int_0^\infty x\mu_X(x)(1-{}_xq_0); \qquad \text{por} \quad (1{,}23)
\end{aligned}
$$

Como $1 - {}_xq_0 = {}_xp_0$,

$$e_0^0 = E[X] = \int_0^\infty x_xp_0\ \mu_X(x)$$

- La esperanza de vida o vida probable para una unidad del grupo en estudio, que ha alcanzado el instante de tiempo y, $(y > 0)$; se denota e_y^0 y está dada mediante la siguiente expresión:

$$e_y^0 = E[T_y] = E[X|X > y] - y \tag{1.31}$$

donde, $T_y = X - y$.

La expresión (1.31), se demuestra de la siguiente manera:

$$
\begin{aligned}
e_y^0 = E[T_y] &= E[X - y]\\
&= E[X] - y \qquad \text{por prop. de esperanza de una variable.}
\end{aligned}
$$

Como la variable X esta condicionada por el evento $X > y$,

$$E[X] = E[X|X > y]$$

Luego,

$$e_y^0 = E[T_y] = E[X|X > y] - y$$

- Para una unidad del grupo en estudio que ha alcanzado el instante de tiempo x, $x > 0$; la esperanza de vida o vida probable de la unidad entre x y $x+1$, es decir, en el intervalo de tiempo $(x,\ x+1)$, está dada mediante la siguiente expresión:

$$E[T_x] = E[T_x|T_x < 1] = \frac{\int_0^1 t\ f_{T_x}(t)\ dt}{\int_0^1 f_{T_x}(t)\ dt} \tag{1.32}$$

Por medio de la expresión (1.30),

$$E[T_x] = E[T_x|T_x < 1] = \frac{\int_0^1 t \; {}_tp_x \; \mu_X(x+t) \; dt}{\int_0^1 {}_tp_x \; \mu_X(x+t) \; dt} \qquad \text{para} \quad 0 < t < 1 \quad (1.33)$$

La siguiente proposición[9] muestra una expresión simple para una variable aleatoria continua con esperanza finita, que permite la demostración del teorema 1.

Proposición 1. Si X es una variable aleatoria continua con esperanza finita, se tiene que:

$$E[X] = \int_0^{+\infty} (1 - F_X(x)) \; dx - \int_{-\infty}^0 F_X(x) dx$$

Teorema 1. *Si T_x es una variable aleatoria continua no negativa para todo $x > 0$, fijo con función de distribución de probabilidad F_{T_x} tal que $F_{T_x}(0) = 0$, y z es una función positiva, diferenciable y monótona con respecto a t, tal que $E[z(T_x)]$ existe, entonces:*

$$E[z(T_x)] = \int_0^{\infty} z(t) f_{T_x}(t) \; dt = z(0) + \int_0^{\infty} z'(t)(1 - F_{T_x}(t)) \; dt$$

Demostración . Como $E[z(T_x)]$ existe

$$E[z(T_x)] = \int_0^{\infty} z(t) f_{T_x}(t) \; dt = \lim_{b \to \infty} \int_0^b z(t) f_{T_x}(t) \; dt$$

Hallemos la integral

$$\int_0^b z(t) f_{T_x}(t) \; dt \qquad (1.34)$$

Integrando por partes la expresión (1.34), y por ser z diferenciable,

[9]Tomada del libro de Probabilidad de Liliana B. Castañeda. Ed. Universidad Nal. Colombia, Bogotá 2004, pág. 203.

$$\int_0^b z(t) f_{T_x}(t)\ dt = F_{T_x}(b) z(b) - z(0) F_{T_x}(0) - \int_0^b z'(t) F_{T_x}(t) dt$$

Como $F_{T_x}(0) = 0$ y además $z(0)$ es constante,

$$\int_0^b z(t) f_{T_x}(t)\ dt = F_{T_x}(b) z(b) - \int_0^b z'(t) F_{T_x}(t) dt$$

Como $S_{T_x}(b) = 1 - F_{T_x}(b)$,

$$\begin{aligned}
\int_0^b z(t) f_{T_x}(t)\ dt &= (1 - S_{T_x}(b)) z(b) - \int_0^b z'(t) F_{T_x}(t) dt \\
&= z(b) - z(b) S_{T_x}(b) - \int_0^b z'(t)(1 - S_{T_x}(t)) dt \\
&= z(b) - z(b) S_{T_x}(b) - \int_0^b z'(t) dt + \int_0^b z'(t) S_{T_x}(t) dt \\
&= z(b) - z(b) S_{T_x}(b) - z(t)|_0^b + \int_0^b z'(t) S_{T_x}(t) dt \\
&= z(0) - z(b) S_{T_x}(b) + \int_0^b z'(t) S_{T_x}(t) dt \qquad (1.35)
\end{aligned}$$

Como z es monótona, consideremos los siguientes casos:

a) Si z es una constante,

De la expresión (1.35),

$$\int_0^b z(t) f_{T_x}(t)\ dt = z(0) - z(b) S_{T_x}(b)$$

como z es constante, $z(b) = z(0)$.

luego,

$$\int_0^b z(t) f_{T_x}(t)\ dt = z(0) - z(0) S_{T_x}(b)$$

por consiguiente,

$$\begin{aligned}
\lim_{b\to\infty}\int_0^b z(t) f_{T_x}(t)\ dt &= \lim_{b\to\infty}(z(0) - z(0)S_{T_x}(b)) \\
&= z(0) - z(0)\lim_{b\to\infty} S_{T_x}(b) \\
&= z(0) - z(0)\cdot 0 \\
&= z(0) - 0 \\
&= z(0)
\end{aligned}$$

vale decir,

$$E[z(T_x)] = z(0) = z(t_0); \qquad \forall t_0 \in \mathbb{R}$$

Por lo tanto,

$$E[z(T_x)] = z(0) + 0 = z(0) + \int_0^\infty z'(t) S_{T_x}(t)\ dt$$

donde, $0 = \int_0^\infty z'(t) S_{T_x}(t)\ dt$

b) Supongamos que z es estrictamente creciente y consideremos la función de distribución de probabilidad de la variable Y, $Y = z(T_x)$. Tal que, para cada $y \in \mathbb{R}$,

$$F_Y(y) = P(Y \leq y) = P(z(T_x) \leq y) = P(T_x \leq z^{-1}(y)) = F_{T_x}(z^{-1}(y))$$

como $Y = z(T_x)$ es una variable no negativa, por la proposición 1.,

$$E[Y] = \int_0^{+\infty} (1 - F_Y(y)) dy$$

Luego,

$$\begin{aligned}
E[Y] &= \int_0^{z(0)} (1 - F_Y(y))dy + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy \\
&= \int_0^{z(0)} dy - \int_0^{z(0)} F_Y(y)dy + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy \\
&= y|_0^{z(0)} - \int_0^{z(0)} F_Y(y)dy + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy \\
&= z(0) - 0 - \int_0^{z(0)} F_Y(y)dy + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy \\
&= z(0) - \int_0^{z(0)} F_Y(y)dy + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy
\end{aligned}$$

Hallemos la integral,

$$\int_0^{z(0)} F_Y(y)dy$$

Como la función z es creciente $z(0) \leq z(t_i)$ Para todo $(z(t_i) > 0)$, lo que conduce a afirmar que $z(0) = 0$. Por consecuencia,

$$\int_0^{z(0)} F_Y(y)dy = F_Y(0) = 0$$

$$E[Y] = z(0) + \int_{z(0)}^{+\infty} (1 - F_Y(y))dy$$

Por último, hallemos la integral,

$$\int_{z(0)}^{+\infty} (1 - F_Y(y))dy$$

Como $Y = z(T_x)$, se tiene que para cada $y \in \mathbb{R}$, $F_Y(y) = F_{T_x}(z^{-1}(y))$.

haciendo una sustitución, $t = z^{-1}(y)$, se obtiene que $y = z(t)$, y $dy = z'(t)dt$,

$$\begin{aligned}\int_{z(0)}^{+\infty}(1-F_Y(y))dy &= \int_{z(0)}^{+\infty}(1-F_{T_x}(t))z'(t)dt \\ &= \int_{z(0)}^{+\infty}z'(t)(1-F_{T_x}(t))dt\end{aligned}$$

como $F_Y(y) = F_{T_x}(z^{-1}(y))$ para cada $y \in \mathbb{R}$,

$$\int_{z(0)}^{+\infty}(1-F_Y(y))dy = \int_{0}^{+\infty}z'(t)(1-F_{T_x}(t))dt$$

Por lo tanto,

$$E[z(T_x)] = z(0) + \int_{0}^{+\infty}(1-F_{T_x}(t))dt$$

Un argumento similar se puede utilizar en el caso en el que es estrictamente decreciente o, en su defecto, es válida la siguiente demostración:

c) Si z es decreciente:

Como z es una función positiva, continua y decreciente, se tiene que z está acotada en $[0, +\infty)$.

Además,

$$\lim_{b\to\infty} S_{T_x}(b) = 0$$

Así que:

$$\lim_{b\to\infty} z(b)S_{T_x}(b) = 0$$

Por lo tanto, de la expresión (1.35) podemos concluir que:

$$E[z(T_x)] = z(0) + \int_{0}^{\infty}z'(t)(1-F_{T_x}(t)dt)$$

$\square$

Proposición 2. Si la esperanza de vida o valor probable de una unidad del grupo en estudio, que ha alcanzado el instante de tiempo y, $y > 0$, esta dada mediante la expresión $e_y^0 = E[T_y] = E[X|X > y] - y$, entonces

$$e_y^0 = \int_0^{\infty} {}_tp_y \; dt$$

Demostración. Como $E[X] = \int_0^{\infty} x f_X(x) \; dx$,

$$e_y^0 = E[X|X > y] - y = \int_0^{\infty} x f_X(x|X > y) dx - y$$

Como T_y es una variable aleatoria continua, de la expresión (1.3),

$$\int_{-\infty}^{+\infty} f_X(x|X > y) dx = 1$$

luego,

$$\begin{aligned} e_y^0 &= \int_0^{\infty} x f_X(x|X > y) dx - y \left(\int_{-\infty}^{+\infty} f_X(x|X > y) dx \right) \\ &= \int_0^{\infty} x f_X(x|X > y) dx - y \left(\int_{-\infty}^{0} f_X(x|X > y) dx + \int_0^{+\infty} f_X(x|X > y) dx \right) \end{aligned}$$

Para $x \leq 0$, $f_X(x|X > y) = 0$, entonces

$$\begin{aligned} e_y^0 &= \int_0^{\infty} x f_X(x|X > y) dx - y \int_0^{+\infty} f_X(x|X > y) dx \\ &= \int_0^{\infty} (x - y) f_X(x|X > y) dx \end{aligned}$$

Haciendo una sustitución, $t = x - y$, $\quad x = t + y \quad$ y $\quad dx = dt$.

$$\begin{aligned} e_y^0 &= \int_0^{\infty} t f_X(t + y|X > y) dt \\ &= \int_0^{\infty} t f_X(t + y) dt \qquad \text{por la observación 1.5.1, numeral 3.} \\ &= \int_0^{\infty} t f_{T_y}(t) dt \end{aligned}$$

por el teorema 1., haciendo $z(t) = t$, donde $z'(t) = 1$,

$$e_y^0 = \int_0^{\infty} t f_{T_y}(t)dt = z(0) + \int_0^{\infty} z'(t)(1 - F_{T_y}(t))dt$$

como $z(0) = 0$ y $z'(t) = 1$

$$\begin{aligned} e_y^0 &= \int_0^{\infty} (1 - F_{T_y}(t))dt \\ &= \int_0^{\infty} (1 - {}_tq_y)dt; \qquad \text{por} \quad (1{,}23) \end{aligned}$$

como $(1 - {}_tq_y) = {}_tp_y$

$$e_y^0 = \int_0^{\infty} {}_tp_y \ dt$$

□

Ejemplo 1.3

Teniendo en cuenta las funciones de supervivencias consideradas en los ejemplos 1.1 y 1.2, calculemos la esperanza de vida o vida probable para las variables X y T_x, respectivamente.

- En el ejemplo 1.1, la función de supervivencia está dada mediante la siguiente expresión:

  $$S_X(x) = 1 - \frac{x}{b}; \qquad \text{para} \quad x \in [0, b].$$

  donde, la variable aleatoria continua X tiene una distribución uniforme en el intervalo [0, b].

De la expresión (1.29), se tiene:

$$
\begin{aligned}
E[X] = e_0^0 &= \int_0^b x f_X(x) dx \\
&= \int_0^b x \left(\frac{1}{b} \right) dx \qquad 0 \leq x < b \\
&= \frac{1}{b} \int_0^b x dx \\
&= \frac{1}{2b} x^2 |_0^b \\
&= \frac{b}{2}
\end{aligned}
$$

Por lo tanto, la cantidad de tiempo esperado de vida o tiempo promedio de vida para el grupo de individuos a quienes se les diagnosticó una cierta enfermedad, esta expresada mediante el valor $\frac{b}{2}$.

- En el ejemplo 1.2, la función de supervivencia está dada mediante la siguiente expresión:

$$S_{T_x}(t) = 1 - \frac{t}{36 - x}; \qquad \text{para} \quad t \in [0,\ 36 - x), \quad x \geq 36 - x.$$

donde, la variable aleatoria continua T_x representa el tiempo de duración de un objeto automático, dado que sale (vende) del almacén después de haber permanecido x tiempo (meses).

De la expresión (1.30), se tiene:

$$E[T_x] = e_x^0 = \int_0^{(36-x)-x} {}_t p_x dt$$

De la expresión (1.23),

$$\begin{aligned}
E[T_x] = e_x^0 &= \int_0^{36-2x} F_{T_x}(t)dt \\
&= \int_0^{36-2x} \frac{t}{36-x} dt \\
&= \frac{1}{36-x} \int_0^{36-2x} tdt \\
&= \frac{1}{36-x} t^2|_0^{36-2x} \\
&= \frac{(36-2x)^2}{2(36-x)}
\end{aligned}$$

1.6.1. El tiempo futuro de vida expresado mediante un número entero.

Teniendo en cuenta la definición de la variable T_x para cada $x > 0$ fijo, se puede obtener para una unidad del grupo en estudio, que ha alcanzado el instante de tiempo x, el tiempo esperado de supervivencia expresado como un número entero[10].

Consideremos Θ el espacio muestral ya definido, y la variable aleatoria K, que representa *el tiempo de supervivencia expresado mediante un número entero para una unidad del grupo en estudio que ha alcanzado el instante de tiempo* x, definida por:

$$K = [T_x] \ ,$$

donde los corchetes representan la función parte entera.

Observación 1.6.1

1. Sea K es una variable aleatoria discreta tal que $K = [T_x]$ para cada $x > 0$ fijo, la función de masa de probabilidad de K, denotada p_K, se obtiene de la siguiente manera:

[10]Evidentemente, se trata de los números enteros positivos, junto con el cero; debido a que los tiempos, particularmente el tiempo de supervivencia se expresa mediante cantidades positivas.

Si $K = [T_x]$ es una variable discreta, por la definición 12.

$$P_K(C) = P(K \in C) = \sum_{k \in C} p_K(k)$$

donde, $C = \mathbb{Z}^+ \cup \{0\}$.

Además, por la condición de normalización

$$P_K(C) = \sum_{k \in C} p_K(k) = \sum_{k \in C} P(K = k) = 1$$

Como T_x es una variable aleatoria continua,

$$P(T_x = k) = P(T_x = k+1) = 0 \quad \text{para} \quad k \geq 0$$

Luego,

$$\begin{aligned}
p_K(k) &= P(k \leq T_x < k+1) \\
&= P(T_x \geq k) - P(T_x \geq k+1) \quad \text{por la observacion 1.2.6, numeral 4.} \\
&= [P(T_x > k) + P(T_x = k)] - [P(T_x > k+1) + P(T_x = k+1)] \\
&= P(T_x > k) - P(T_x > k+1) \\
&= {}_k p_x - {}_{k+1} p_x
\end{aligned}$$

Como ${}_k p_x = 1 - {}_k q_x$, y, $\quad {}_{k+1} p_x = 1 - {}_{k+1} q_x$

$$p_K(k) = (1 - {}_k q_x) - (1 - {}_{k+1} q_x) = {}_{k+1} q_x - {}_k q_x$$

Por la observación (1.5.1), numeral 2,

$$p_K(k) = {}_{k|1} q_x \tag{1.36}$$

2. Con base en la expresión anterior, y teniendo en cuenta la expresión (1.9), la función de distribución de probabilidad de la variable K, denotada F_K, está dada de la siguiente manera:

$$F_K(k) = P(K \leq k) = \sum_{k \in C} p_K(k) = \sum_{h=0}^{k} {}_{h/1} q_x \tag{1.37}$$

donde, $C \in \mathbb{R}$, y es un conjunto finito o infinito numerable donde los valores de $p_K(k)$ son positivos.

Nótese que,

$$\sum_{h=0}^{k} {}_{h/1}q_x = {}_{k+1}q_x = \begin{cases} 0, & \text{si} \quad k < 0 \\ P(T_x \leq k+1), & \text{si} \quad k \geq 0 \\ 1, & \text{si} \quad k > \omega \end{cases} \tag{1.38}$$

3. Como K es una variable aleatoria discreta, por la expresión (1.14), la esperanza de vida o vida probable para una unidad del grupo en estudio que ha alcanzado el instante de tiempo x, $(x > 0)$, se denota e_x, está dada de la siguiente manera:

$$e_x = E[K] = \sum_{k=0}^{\infty} k \ p_K(k) = \sum_{k=0}^{\infty} k \ ({}_{k/1}q_x)$$

Por la observación 1.5.1, numeral 2: ${}_{k/1}q_x = {}_kp_x \cdot {}_1q_{x+k}$

Luego,

$$e_x = E[K] = \sum_{k=0}^{\infty} k \ {}_kp_x \cdot {}_1q_{x+k} \tag{1.39}$$

Nota: *La esperanza de vida o vida probable en el caso discreto para una unidad del grupo en estudio, que ha alcanzado el instante de tiempo x, se denota e_x, para diferenciarla de la esperanza de vida en el caso continuo, e_x^0; debido a que más adelante se necesitará una distinción entre ellas; pero esto no quiere decir que no pueda notarse como e_x^0.*

La siguiente proposición, es útil para la demostración del teorema 2, que presenta una expresión simple para obtener la esperanza de vida o vida probable de una variable discreta.

Proposición 3. Si X es una variable aleatoria discreta y no negativa con esperanza finita entonces

$$E[X] = \sum_{x=0}^{\infty} xP(X = x)$$

Teorema 2. *Como K es una variable aleatoria discreta con función de distribución F_K y función de masa de probabilidad p_K, tal que $p_K(k) = \triangle\ p_K(k-1) = p_K(k) - p_K(k-1)$, y z una función positiva y monótona, tal que $E[z(K)]$ existe, entonces:*

$$E[z(K)] = \sum_{k=0}^{\infty} z(k)p_K(k) = z(0) + \sum_{k=0}^{\infty}(1 - F_K(k))\triangle z(k)$$

donde, $\triangle z(k) = z(k+1) - z(k)$ para $k = 0, 1, \ldots$

Demostración. como $E[z[K]]$ existe,

$$E[z[K]] = \sum_{k=0}^{\infty} z(k)p_K(k) = \lim_{n\to\infty} \sum_{k=0}^{n} z(k)p_K(k)$$

Calculemos,

$$\sum_{k=0}^{n} z(k)p_K(k)$$

Como,

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0)p_K(0) + \sum_{k=1}^{n} z(k)p_K(k)$$

Como K es una variable discreta y F_K es la función de distribución de probabilidad de la variable K, se tiene que

$$p_K(k) = F_K(k) - F_K(k-1) \quad \text{para} \quad k = 0, 1, \cdots \quad \text{donde} \quad F_K(-1) = 0$$

Luego,

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0)[F_K(0) - F_K(0-1)] + \sum_{k=1}^{n} z(k)[F_K(k) - F_K(k-1)]$$

Como $F_K(0) = 1$ y $F_K(-1) = 0$

$$\begin{aligned}\sum_{k=0}^{n} z(k)p_K(k) &= z(0) + \sum_{k=1}^{n} z(k)[F_K(k) - F_K(k-1)] \\ &= z(0) + \sum_{k=1}^{n} z(k)F_K(k) + \sum_{k=1}^{n} z(k)F_K(k-1)\end{aligned}$$

Como $\sum_{k=1}^{n} z(k)F_K(k-1) = \sum_{m=0}^{n-1} z(m+1)F_K(m)$

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) + \sum_{k=1}^{n} z(k)F_K(k) - \sum_{m=0}^{n-1} z(m+1)F_K(m)$$
$$= z(0) + \left[\sum_{k=1}^{n-1} z(k)F_K(k) + z(n)F_K(n)\right] - \left[z(1)F_K(0) + \sum_{m=1}^{n-1} z(m+1)F_K(m)\right]$$

Haciendo $k = m$

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) + \sum_{k=1}^{n-1} F_K(k)[z(k) - z(k+1)] - z(1)F_K(0) + z(n)F_K(n)$$
$$= z(0) - \sum_{k=1}^{n-1} F_K(k)[z(k+1) - z(k)] + z(n)F_K(n) - z(1)F_K(0)$$

Sumando y restando a un lado de la igualdad $\sum_{k=1}^{n-1}[z(k+1) - z(k)]$

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) - \sum_{k=1}^{n-1} F_K(k)[z(k+1) - z(k)] + z(n)F_K(n) - z(1)F_K(0) + A$$

donde,

$$A = \sum_{k=1}^{n-1}[z(k+1) - z(k)] - \sum_{k=1}^{n-1}[z(k+1) - z(k)]$$

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) + \sum_{k=1}^{n-1}[z(k+1) - z(k)](1 - F_K(k)) + z(n)F_K(n)$$
$$- z(1)F_K(0) - B$$

donde,

$$B = \sum_{k=1}^{n-1} [z(k+1) - z(k)]$$

Calculemos,

$$\sum_{k=1}^{n-1} [z(k+1) - z(k)] = z(n) - z(1)$$

Por consiguiente,

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)F_K(n)$$
$$-z(1)F_K(0) - [z(n) - z(1)]$$
$$= z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)F_K(n) -$$
$$z(1)F_K(0) - z(n) + z(1)$$
$$= z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)F_K(n) -$$
$$z(n) + z(1)(1 - F_K(0))$$

Como $F_K(0) = 1$,

$$\begin{aligned}\sum_{k=0}^{n} z(k)p_K(k) &= z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)F_K(n) - z(n) \\ &= z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)(F_K(n) - 1)\end{aligned}$$

Por tanto,

$$\sum_{k=0}^{n} z(k)p_K(k) = z(0) + \sum_{k=1}^{n-1} [z(k+1) - z(k)](1 - F_K(k)) + z(n)(F_K(n) - 1) \tag{1.40}$$

Como z es una función monótona, consideremos los siguientes casos:

- Si z es decreciente:

  Como z es una función positiva, continua y decreciente, se tiene que z esta acotada en $[0, +\infty)$.

  Además,
  $$\lim_{n\to\infty} [F_K(n) - 1] = 0$$

  Entonces,
  $$\lim_{n\to\infty} z(n)(F_K(n) - 1) = 0$$

  Por lo tanto,
  $$\begin{aligned} \lim_{n\to\infty} \sum_{k=0}^{n} z(k)p_K(k) &= \lim_{n\to\infty} z(0) + \sum_{k=1}^{n-1}[z(k+1) - z(k)](1 - F_K(k)) \\ &= z(0) + \sum_{k=0}^{\infty} \triangle z(k)(1 - F_K(k)) \end{aligned}$$

- Si z es creciente:

  Teniendo en cuenta la proposición 3., y haciendo un racionamiento similar al realizado en el teorema 1., se comprueba que z es creciente.

Por lo tanto, de la expresión (1.40) podemos concluir que:

$$E[z(K)] = z(0) + \sum_{k=0}^{\infty}(1 - F_K(k))\triangle z(k)$$

□

Proposición 4. Si la esperanza de vida o vida probable, en el caso discreto, para una unidad del grupo en estudio, que ha alcanzado el instante de tiempo x, $(x > 0)$ esta dada por $e_x = \sum_{k=0}^{\infty} k \; {}_kp_x \cdot {}_1q_{x+k}$ entonces:

$$e_x = \sum_{k=1}^{\infty} {}_kp_x$$

Demostración. Por hipótesis

$$e_x = \sum_{k=0}^{\infty} k \ {}_k p_x \cdot {}_1 q_{x+k}$$

Por la observación 1.5.1, numeral 2b,

$$e_x = \sum_{k=0}^{\infty} k_{k|1} q_x$$

De la expresión (1.36),

$$p_K(k) = {}_{k|1} q_x$$

luego,

$$e_x = \sum_{k=0}^{\infty} k p_K(k)$$

donde, p_K es la función de masa de probabilidad de la variable discreta K.

Tomando $z(k)$ como la función idéntica $z(k) = k$

$$e_x = \sum_{k=0}^{\infty} z(k) p_K(k)$$

Por el teorema 2,

$$\begin{aligned} e_x = \sum_{k=0}^{\infty} z(k) p_K(k) &= z(0) + \sum_{k=0}^{\infty} (1 - F_K(k)) \triangle z(k) \\ &= 0 + \sum_{k=0}^{\infty} (1 - F_K(k)) \triangle z(k) \end{aligned}$$

Como $\triangle z(k) = z(k) - z(k-1) = k - (k-1) = 1$

Por lo tanto,

$$e_x = \sum_{k=0}^{\infty}(1 - F_K(k))$$

Como $K = [T_x]$, de la expresión (1.23),

$$\begin{aligned} e_x &= \sum_{k=0}^{\infty}(1 - {}_{(k+1)}q_x) \\ &= \sum_{k=1}^{\infty}(1 - {}_kq_x) \end{aligned}$$

Como $1 - {}_kq_x = {}_kp_x$

$$e_x = \sum_{k=1}^{\infty} {}_kp_x$$

$\square$

Observación 1.6.2

Si se desea calcular la fracción[11] de tiempo de supervivencia en el intervalo de tiempo en el que se experimenta el evento de interés, consideremos el espacio muestral Θ ya definido, y sea $h, \quad 0 \leq h < 1$ la variable que representa *la fracción de tiempo de supervivencia en el intervalo de tiempo en el que se experimenta el evento de interés.* Para una unidad del grupo en estudio que ha alcanzado el instante de tiempo x, tal que para cada $x > 0$, fijo:

$$h = T_x - K.$$

Esto quiere decir, que si la unidad del grupo en estudio ha alcanzado el instante de tiempo x y logra permanecer en el estudio k tiempo más, es decir, logra alcanzar el instante de tiempo $x + k$, pero experimenta el evento de interés entre $x + k$ y $x + k + h$ con $0 \leq h < 1$, entonces:

[11] Indica una parte o porción pequeña de alguna cosa. La fracción de tiempo en el acontecimiento de un evento es el último instante de tiempo que se observa antes de que ocurrir el evento esperado.

$$P(k < T_x \leq k+h) = P(T_x > k) - P(T_x > k+h)$$

Esto es, por la observación 1.2.6

Luego,

$$P(k < T_x \leq k+h) = {}_kp_x \cdot {}_hq_{x+k} \qquad 0 \leq h < 1$$

Gran parte de los análisis estadísticos se basan en la hipótesis de que la muestra de tiempos observados se distribuye según una normalidad. En este caso el estudio se reduce a la estimación de un número finito de cantidades (parámetros), a saber, la media y al varianza de la distribución en el caso normal. Sin embargo, este método no se adapta lo suficientemente bien a muestras formadas por tiempos de supervivencia. Los métodos paramétricos, es decir, aquellos que no presuponen que los datos se ajusten a una distribución concreta, son muy útiles en el análisis de supervivencia.

Capítulo 2

Algunos métodos estadísticos en el análisis de Supervivencia.

2.1. Introducción

Desde ya hace algunos años, en diferentes ramas como la industria, la medicina, entre otras, para el ánalisis de tiempos de supervivencia y de riesgo, se han utilizado algunos métodos estadísticos no-paramétricos como el método de la Tabla de vida, el método de kaplan-Meier para estimar e interpretar a través de la función de supervivencia obtenida por cada método, la descripción y resumen de los datos, los cuales no han sido ajustados a una distribución en particular. Mediante la comparación de las distribuciones de los tiempos de supervivencia correspondientes a dos o ms poblaciones diferentes y el establecimiento y comprensión de la relación que pueda existir entre los tiempos y las variables explicativas puede realizarse el método estadístico no-paramétrico de Log-Rank.

En éste documento, se establece una descripción de las características propias de los métodos estadísticos no- paramétricos: método de la tabla de vida, método de Kaplan-Meier, y el método de Log-Rank para la estimación de la función de supervivencia.

2.2. El Método de la Tabla de Vida.

El método clásico para determinar la curva de supervivencia en una muestra de observaciones tomada de la población en estudio, es la denominada “Tabla de Vida”, la cual es una técnica que explora métodos de inferencia estadística para calcular la función de supervivencia en intervalos de tiempo. ..... Los tiempos de supervivencia se agrupan en intervalos, donde la longitud del intervalo depende de la frecuencia con que se presenta el evento de interés (no necesariamente, los intervalos son de una misma longitud). Cuando se emplea éste método para el análisis de datos de supervivencia o de riesgo, los valores de la función de supervivencia, cambia sólo durante los períodos de tiempo que contienen al menos un instante de tiempo donde alguna unidad del estudio presentó el evento de interés.

Una tabla de vida, contiene distintas informaciones descriptivas sobre la evolución de las observaciones. Para la construcción de una tabla de vida, se procede de la siguiente manera:

- La **Columna 1**: Contiene los tiempos de supervivencia, agrupados en intervalos de amplitud n, $(n > 0)$ fija o variada.
- En la **Columna 2**: Se relaciona la proporción de unidades que experimentaron el evento de interés durante el intervalo de tiempo $(t,\ t+n)$, dado que las unidades habían alcanzado el instante de tiempo t, esto es ${}_nq_t$.
- La **Columna 3**: Contiene el número esperado de unidades que no han experimentado el evento de interés, en cada instante de tiempo t, denotado l_t.
- En la **Columna 4**: Se relaciona la proporción de unidades que aún no han experimentado el evento de interés en el instante de tiempo $t+n$, dado que las unidades habían alcanzado el instante de tiempo t; esto es ${}_np_t$.
- La **Columna 5**: Contiene la proporción acumulada de unidades en cada intervalo de tiempo; esto es, la función de supervivencia evaluada en cada instante de tiempo t. Esta proporción acumulada puede estimarse de las siguientes maneras:

  1. Si l_0 denota la cantidad de unidades en el estudio, que no experimentaron el evento de interés en instante de tiempo inicial $t = 0$, y l_t el número esperado de unidades del estudio que no han experimentado el evento de interés en el instante de tiempo t. La proporción de unidades que no han experimentado el evento de interés en el instante de tiempo t, puede determinarse mediante la siguiente expresión:

$$S_X(t) = \frac{l_t}{l_0} \tag{2.1}$$

2. De la expresión (1.16), se tiene que dados dos eventos A y B,

$$P(A \cap B) = P(A)P(B|A) \qquad \text{si} \quad P(A) > 0$$

A través la expresión anterior, puede calcularse la función de supervivencia S_X en un intervalo de tiempo. Consideremos los eventos A y B definidos de la siguiente forma:

A : Sea el evento de que una unidad del estudio no experimente el evento de interés durante el intervalo de tiempo de 0 a 1 mes: $(0, 1)$, y

B : Sea el evento de que una unidad del estudio no experimente el evento de interés hasta el instante de tiempo 1 mes.

Por tanto, el evento de que la unidad no experimente el evento de interés más alla de un 1 mes, puede representarse por $A \cap B$. Por (1.16), la probabilidad de que la unidad no experimente el evento de interés más alla de un 1 mes, es $P(X > 1) = S_X(1)$, y la probabilidad de que la unidad no experimente el evento de interés durante el intervalo de tiempo $(0, 1)$ es $S_X(0)$. Ademas, la probabilidad de que la unidad no experimente el evento de interés hasta el instante de tiempo de un 1 mes, dado que la unidad no lo ha experimentado hasta ese instante, es decir, $(1 - {}_1q_1) = {}_1p_1$. Entonces,

$$S_X(1) = S_X(0) * {}_1p_1$$

Nota: *Una tabla de vida puede contener más columnas, todo depende de la cantidad de información necesaria para determinar una curva de supervivencia.*

Ejemplo 2.1

El siguiente ejemplo fue tomado del autor Alejandro Predomingo, en su obra "Fundamentos Teóricos del Análisis de Sobrevivencia", pág. 492.

La siguiente tabla muestra los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte para una muestra de 12 pacientes del hospital central de Guatemala, de a lo sumo 40 años de edad en el momento del diagnostico.

TABLA 2.1

Número del paciente	Supervivencia(meses)
1	2
2	3
3	6
4	6
5	7
6	10
7	15
8	15
9	16
10	27
11	30
12	32

Observando la tabla 2.1 y teniendo en cuenta que un tiempo de supervivencia de t meses, significa que un paciente de la muestra tomada de la población en estudio, sobrevivió t, desde el momento en el que se le diagnóstico SIDA hasta el instante de tiempo en el que murió. Seguiendo el procedimiento de la tabla de vida descrito antes, se tiene el siguiente análisis:

- En la **Columna 1**: Se dividen los tiempos de supervivencia en intervalos de amplitud fija de un (1) mes. Esto es, por que no existe una dispersión amplia entre los tiempos de supervivencia.

- La **Columna 2**: Contiene la proporción de pacientes que murieron durante el intervalo de tiempo $(t,\ t+1)$, dado que habían sobrevivido al instante de tiempo t.

  - Durante los intervalos de tiempo $(0,1)$ mes y $(1,2)$ mes, no se registró ninguna muerte; esto quiere decir:

$$
\begin{aligned}
{}_1q_0 &= \frac{0}{12} = 0 \\
{}_1q_1 &= \frac{0}{12} = 0
\end{aligned}
$$

  - Como puede verse en la tabla 2.1, uno (1) de los 12 pacientes del grupo en estudio murió a los dos(2) meses después de haberle diagnosticado SIDA. En una muestra de 12 pacientes, la proporción de pacientes que murieron en el intervalo de tiempo $(2, 3)$ meses se calcula de la siguiente manera:

$$_1q_2 = \frac{1}{12} = 0{,}0833$$

- Uno (1) de los 11 pacientes restantes del grupo inicial, murió a los 3 meses después de haberle diagnosticado SIDA. La proporción de pacientes que murieron en el intervalo de tiempo $(3, 4)$ meses se calcula de la siguiente manera::

$$_1q_3 = \frac{1}{11} = 0{,}0909$$

- La **Columna 3**: Contiene el número esperado de sobrevivientes en cada intervalo de tiempo.

  - Durante el intervalo de tiempo $2-3$ meses, un 1 individuo murió; esto quiere decir que el número esperado de sobrevivientes para el intervalo de tiempo $3-4$ meses es de 11 individuos.

- La **Columna 4**: Contiene la proporción de pacientes que sobrevivieron al instante de tiempo $t+1$, dado que habían sobrevivido al instante de tiempo t.

  - Como durante los intervalos de tiempo $0-1$ mes y $1-2$ meses no se registraron muertes, la proporción de pacientes vivos en estos intervalos es de $1, 0$; es decir: $_1p_0 = 1$ y $_1p_1 = 1$
  - En el intervalo de tiempo $2-3$ meses, 11 pacientes sobrevivieron al instante de tiempo 3 meses, dado que 12 pacientes habían sobrevivido al instante de tiempo 2 meses:

$$_1p_2 = 1 - {_1q_2} = 1 - 0{,}0833 = 0{,}9167$$

- En la **Columna 5**: Aparece el valor de la función de supervivencia evaluada en cada instante de tiempo t:

  - Ningún individuo en la muestra murió en el instante de tiempo (0) meses. La probabilidad de que un individuo cualquiera sobreviva durante el intervalo de tiempo $0-1$ mes, es:

$$S_X(0) = P(X > 0) = 1{,}0000$$

- La función de supervivencia evaluada en el instante de tiempo un (1) mes, se obtiene de la siguiente manera:

$$S_X(1) = S_X(0)[{}_1p_1] = (1{,}0000)(1{,}0000) = 1{,}0000$$

- La función de supervivencia evaluada en el instante de tiempo 2 meses, se obtiene de la siguiente manera:

$$S_X(2) = S_X(1)[{}_1p_2] = (1{,}0000)(0{,}9167) = 0{,}9167$$

Como puede verse en la tabla 2.1, después de 32 meses de supervivencia, ninguno de los 12 pacientes de la muestra se encuentra vivo; como resultado $S_X(32) = 0$.

La aplicación de este método al ejemplo anterior, nos permite construir la siguiente tabla:

TABLA 2.2

Tabla de vida para calcular la función de supervivencia de los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes del hospital central de Guatemala de a lo sumo 40 años de edad en el momento del diagnostico.

t a t+n	${}_nq_t$	l_t	${}_np_t$	$S_X(t)$
0-1	0.0000	12	1.0000	1.0000
1-2	0.0000	12	1.0000	1.0000
2-3	0.0833	12	0.9167	0.9167
3-4	0.0909	11	0.9091	0.8333
4-5	0.0000	10	1.0000	0.8333
5-6	0.0000	10	1.0000	0.8333
6-7	0.2000	10	0.8000	0.6667
7-8	0.1250	8	0.8750	0.5833
8-9	0.0000	7	1.0000	0.5833
9-10	0.0000	7	1.0000	0.5833
10-11	0.1429	7	0.8571	0.5000

11-12	0.0000	6	1.0000	0.5000
12-13	0.0000	6	1.0000	0.5000
13-14	0.0000	6	1.0000	0.5000
14-15	0.0000	6	1.0000	0.5000
15-16	0.3333	6	0.6667	0.3333
16-17	0.2500	4	0.7500	0.2500
17-18	0.0000	3	1.0000	0.2500
18-19	0.0000	3	1.0000	0.2500
19-20	0.0000	3	1.0000	0.2500
20-21	0.0000	3	1.0000	0.2500
21-22	0.0000	3	1.0000	0.2500
22-23	0.0000	3	1.0000	0.2500
23-24	0.0000	3	1.0000	0.2500
24-25	0.0000	3	1.0000	0.2500
25-26	0.0000	3	1.0000	0.2500
26-27	0.0000	3	1.0000	0.2500
27-28	0.3333	3	0.6667	0.1667
28-29	0.0000	2	1.0000	0.1667
29-30	0.0000	2	1.0000	0.1667
30-31	0.5000	2	0.5000	0.0833
31-32	0.0000	1	1.0000	0.0833
32-33	1.0000	1	0.0000	0.0000

Una curva de supervivencia puede aproximarse al trazar una gráfica de la función de supervivencia S_X, generada al aplicar el método de la tabla de vida en función del punto que representa el inicio de cada intervalo, y uniendo los puntos consecutivos con líneas rectas.

GRÁFICA 2.1

Curva de supervivencia de los tiempos de supervivencia de 12 pacientes del hospital central de Guatemala desde el diagnóstico de SIDA hasta la muerte de a lo sumo 40 años de edad en el momento del diagnostico.

GRÁFICA 2.1

Fuente:PREDOMINGO, Alejandro. Fundamentos Teóricos del Análisis de Sobrevivencia, Pág. 493.

Observación 2.2.1

Es de suma importancia para las compañias de seguros o pensiones, como para los estudios de medicina las tablas de mortalidad, las cuales son tablas de vida que resume las experiencias de supervivencia y de mortalidad de un grupo de individuos donde se considera importante el tiempo vivido de cada individuo antes de su muerte.

En la construcción de una tabla de mortalidad para un grupo de individuos se debe tener en cuenta que existen dos formas de tabla de vida: de Cohorte y Actuarial.

- La tabla de cohorte.

  Es una tabla en la se muestra el seguimiento que se le realiza a un grupo real de individuos a lo largo de sus vidas hasta su completa extinción o bien cuando se decide concluir el período de observación. Este proceso resulta poco práctico, porque presenta una serie de dificultades para utilizarla en la descripción de la supervivencia de la población general. La tabla de cohorte se usa habitualmente en el análisis de supervivencia de los ensayos clínicos, que se realizan sobre muestras de población pequeñas y durante un tiempo corto.

- La tabla Actuarial.

  Es una tabla que se construye a partir de los datos recopilados de las experiencias de mortalidad y supervivencia en todas las edades de una población durante un corto período de tiempo, habitualmente un año. En la tabla actuarial se utiliza la experiencia de mortalidad de una población durante cada año determinado, que se aplica a una cohorte ficticia de 10.000 o 100.000 individuos vivos. Aunque el cálculo se basa en una parte "ficticia" (el tamaño de población), la tabla actuarial refleja la experiencia de mortalidad "real" de la población considerada y es una herramienta sumamente útil para comparar datos de mortalidad a nivel internacional y para valorar las tendencias de mortalidad a nivel nacional. La tabla actuarial se usa en problemas relacionados con seguros de vida o pensiones.

Consideremos un grupo (inicial) hipotético de l_0 individuos recién nacidos todos el mismo día o con una edad particular x (preferiblemente $x = 0$), y un tiempo míximo de supervivencia al cual no se puede sobrevivir ω.

El grupo hipotético compuesto por l_0 individuos recién nacidos puede ser considerado *Grupo Aleatorio* si cada vida del grupo tiene una función especifíca de supervivencia; o también *Grupo Determinístico de supervivientes* si después del inicio no hay ingresos, y el número de individuos que conforman el grupo decrece únicamente por muerte.

Definición 16. *Consideremos un grupo inicial compuesto por l_0 individuos recién nacidos y $\mathcal{L}_x : \Omega \longrightarrow \mathbb{R}$ la variable aleatoria que denota(cuenta) el número de sobrevivientes en el tiempo de supervivencia x del grupo y, $I_{j,x}$ la variable de tipo Bernoulli, que representa la supervivencia del j-ésimo individuo del grupo en el instante de tiempo x; tal que para $x > 0$ fijo, se define:*

$$I_{j,x} = \begin{cases} 1, & \textit{si j-ésimo individuo sobrevive a la edad } x \\ 0, & \textit{en otro caso} \end{cases}$$

Como $\mathcal{L}_x$ denota el número de sobrevivientes del grupo en el tiempo de supervivencia x, esta puede expresarse así:

$$\mathcal{L}_x = \sum_{j=1}^{l_0} I_{j,x}$$

Si el número inicial de sobrevivientes es $l_0 = n$, y $S_X(x)$ la proporción de individuos vivos en el instante de tiempo x, entonces la variable aleatoria $\mathcal{L}_x$ se distribuye binomialmente con parámetros l_0 y $S_X(x)$

$$\mathcal{L}_x \sim B(l_0, S_X(x))$$

Luego, ***El número esperado de sobrevivientes l_x en el instante de tiempo*** *x, se obtiene de la siguiente manera:*

$$E[\mathcal{L}_x] = l_x = l_0 S_X(x) \tag{2.2}$$

Con base en la ecuación (2.2) y considerando la variable ${}_kD_x : \Omega \longrightarrow \mathbb{R}$ que denota *el número de muertes en el grupo entre los instantes de tiempo x y $x+k$*, tenemos que:

$${}_kD_x \sim B(l_0, F_X(x+k) - F_X(x))$$

Luego, **El número esperado de muertes ${}_kd_x$ del grupo entre los instantes de tiempos x y $x+k$**, se obtiene de la siguiente manera:

$$E[{}_kD_x] = {}_kd_x = l_0(S_X(x) - S_X(x+k)) = l_x - l_{x+k} \tag{2.3}$$

A partir de las fórmulas (2.2) y (2.3) podemos obtener algunas funciones que hacen parte de la tabla[1] de mortalidad de un grupo hipotético conformado por 100.000 individuos de edad inicial 20 años.

Para las probabilidades condicionales de supervivencia en un grupo, se tiene que:

$${}_1p_x = \frac{l_{x+1}}{l_x} \tag{2.4}$$

Esta fórmula se comprueba de la siguiente manera:

Como ${}_1p_x = \frac{S_X(x+1)}{S_X(x)}$ y $S_X(x) = \frac{l_x}{l_0}$, entonces

$${}_1p_x = \frac{S_X(x+1)}{S_X(x)} = \frac{\frac{l_{x+1}}{l_0}}{\frac{l_x}{l_0}} = \frac{l_{x+1}}{l_x}$$

[1]Anexo A.

Además, se tiene que

$$ {}_1q_x = \frac{{}_1d_x}{l_x} \tag{2.5} $$

Esta fórmula se comprueba de la siguiente manera:

Como ${}_1q_x = \frac{S_X(x)-S_X(x+1)}{S_X(x)}$ y ${}_1d_x = l_x - l_{x+1}$, entonces

$$ {}_1q_x = \frac{S_X(x) - S_X(x+1)}{S_X(x)} = \frac{l_x - l_{x+1}}{l_x} = \frac{{}_1d_x}{l_x} $$

Observación 2.2.2

Comunmente en las tablas de mortalidad las probabilidades condicionales de supervivencia ${}_tq_x$, ${}_tp_x$, ${}_td_x$ aparecen denotadas como: p_x, q_x, d_x indicando que la diferencia entre los tiempos de supervivencia (edades) es de un (1) año. Esto es, ${}_1p_x = p_x$; ${}_1q_x = q_x$; ${}_1d_x = d_x$.

Ejemplo 2.2

Teniendo en cuenta la Tabla[2] Colombiana de Mortalidad de los asegurados de una compañia de seguros durante el período de 1.984 - 1.989, tomada del libro de Huertas Jaime Abel, Continguencias de vida individual: Cálculo Actuarial. Pág. . El grupo inicial está conformado por 100.000 individuos con edad 20 años, interpretemos algunos resultados:

Después de transcurridos cuatro(4) años desde la entrada de los 100.000 individuos con edad inicial 20 años, se tiene:

- El número esperado de sobrevivientes, cuando los individuos han alcanzado los 24 años es de 98.617 individuos, es decir, se espera que el 98,62 % del grupo inicial sobreviva; mientrás que se espera que el 1,38 % del grupo inicial muera en el transcurso de los 4 años.

- Durante el transcurso de los 24 años, el número esperado de muertes a esa edad es de 348 individuos, es decir la proporción de individuos que muere a esa edad es de 0,3529.

- Del grupo de individuos sobrevivientes a los 24 años, se espera que sobrevivan en promedio 48 años más y 9 meses.

[2]ver anexo A.

2.3. Método de Kaplan - Meier

El método de Kaplan - Meier, permite el análisis de datos censurados y no censurados de supervivencia o de riesgo de muestras conformadas por un número grande de observaciones, esto es, por el hecho de calcular en la función de supervivencia obtenida de éste método, los tiempos en los cuales las unidades de la muestra en estudio, experimentaron el evento de interés; es decir, se utilizan los tiempos de supervivencia exactos en que cada unidad del estudio experimentó el evento de interés, en lugar de agrupar estos tiempos en intervalos, como se realiza mediante el método de la tabla de vida. Cuando se aplica el método de kaplan-Meier,la función de supervivencia, S_X, se supone que continúa siendo la misma durante los períodos de tiempo entre los cuales no hay evidencia de que las unidades en el estudio experimentaron el evento de interés; cambia precisamente cuando hay presencia de que al menos una unidad experimentó el evento de interés del estudio.

Para el análisis de datos de supervivencia o de riesgo, es más preciso utilizar el método de Kaplan - Meier, en lugar de las tablas de vida y no está restringido a pequeños grupos de unidades. Además, su procesamiento se puede hacer con ayuda de diferentes paquetes estadísticos.

2.3.1. Curva de supervivencia Kaplan - Meier.

Para determinar una curva de supervivencia por el método de Kaplan-Meier, se deben tener en cuenta las siguientes informaciones descriptivas sobre la evolución de las observaciones tomadas de una población, y se calcula en la función de supervivencia los tiempos en los cuales las unidades del estudio experimentaron el evento de interés:

- La **Columna 1**: Contiene los tiempos exactos en los cuales las unidades del estudio experimentaron el evento de interés. Estos tiempos se ordenan de menor a mayor (censurados y no censurados).

  Nota: *Cuando se tienen observaciones censuradas se coloca junto a ella un signo positivo. Para las observaciones censuradas y no censuradas que tienen el mismo tiempo de supervivencia, se debe colocar primero la observación no censurada.*

- En la **Columna 2**: Se relaciona la proporción de unidades que no han experimentado el evento de interés antes de alcanzar el instante de tiempo

t, con relación a las unidades que experimentaron el evento de interés en el instante de tiempo t. Esto se denota mediante la expresión q_t.

Nótese que la expresión q_t depende unicamente del valor t; a diferencia de ${}_tq_x$ que esta condiciona por el instante de tiempo x.

- En la **Columna 3**: Se relaciona la proporción de unidades que no han experimentado el evento de interés del estudio después de haber alcanzado el instante de tiempo t. Esto se denota, p_t.
- **Columna 4**: Se calcula la función de supervivencia en cada instante de tiempo t, en el que las unidades del grupo en estudio experimentaron el evento de interés.

Notas:

1. *Los tiempos censurados son algunas veces indicados por marcas sobre la curva de supervivencia, los cuales nos facilitan la búsqueda de los tiempos de supervivencia de los sujetos a los cuales no les ha ocurrido el evento de interés.*
2. *En las figuras de curvas de supervivencia, la presencia de líneas horizontales largas, indican que no hay cambios en la probabilidad de supervivencia estimada para un período. A veces esto ocurre porque la muestra es muy pequeña.*

2.3.2. Función de supervivencia calculada por el método Kaplan - Meier.

La Supervivencia en el i-ésimo instante de tiempo t_i; se calcula mediante la siguiente expresión:

$$S_X(t_i) = \frac{r_i - m_i}{r_i} * S_X(t_{i-1}) \tag{2.6}$$

donde,

r_i: es el número de unidades que permanecian (expuestas al riesgo) en el estudio, en el instante anterior, (t_{i-1}).

m_i: El número de unidades que presentaron el evento de interés en el i-ésimo instante de tiempo.

Nótese que para cada instante de tiempo t_i, la supervivencia se calcula como la supervivencia en el instante de tiempo anterior $t_{(i-1)}$ múltiplicada por la tasa de supervivencia en el instante de tiempo t_i.

Ejemplo 2.3

Retomando la información de la tabla de vida 2.2, la estimación de la función de supervivencia mediante el método de Kaplan- Meier, se obtiene se la siguiente manera:

- La **Columna 1**: Contiene los tiempos exactos en que ocurrieron las muertes de los 12 pacientes de a lo sumo 40 años de edad en el momento del diagnostico. Estos son: $0, 2, 3, 6, 7, 10, 15, 16, 27, 30$ y 32.

- En la **Columna 2**: Se relaciona la proporción de pacientes que estaban vivos antes de cada instante de tiempo t, con relación a los que murieron en el instante de tiempo t:

  - Ninguno de los 12 pacientes, murió en el instante de tiempo 0.

  $$q_0 = \frac{0}{12} = 0{,}0000$$

  - 1 de los 12 pacientes, murió en el instante de tiempo 2.

  $$q_2 = \frac{1}{12} = 0{,}0833$$

  - 1 de los 11 pacientes aún vivos, murió en el instante de tiempo 3.

  $$q_3 = \frac{1}{11} = 0{,}0909$$

- En la **Columna 3**: Se relaciona la proporción de pacientes que siguen vivos después del instante de tiempo t:

  - Todos los pacientes se encuentran vivos en el instante de tiempo 0.

  $$p_0 = \frac{12}{12} = 1{,}0000$$

  - 11 de 12 pacientes aún se encuentran vivos después del instante de tiempo 2.

  $$p_2 = \frac{11}{12} = 0{,}9167$$

- 10 de 11 pacientes aún se encuentran vivos después del instante de tiempo 3.

$$p_3 = \frac{10}{11} = 0{,}9091$$

- **Columna 4**: Calculemos los valores de la función de supervivencia en cada instante de tiempo t, teniendo en cuenta la expresión (2.2):

  - La proporción de pacientes que aún no fallecen en el instante de tiempo 0, es de:

$$S_X(0) = P(X > 0) = 1{,}0000$$

  - La proporción de pacientes que aún no fallecen en el instante de tiempo 2:

$$S_X(2) = \frac{12-1}{12} * S_X(0) = 0{,}9167 * 1{,}0000 = 0{,}9167$$

  - La proporción de pacientes que aún no fallecen en el instante de tiempo 3:

$$S_X(3) = \frac{11-1}{11} * S_X(2) = 0{,}9091 * 0{,}9167 = 0{,}8333$$

La tabla 2.3 presenta el cálculo de la función de supervivencia mediante el método de Kaplan- Meier de los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes del hospital central de Guatemala de a lo sumo 40 años de edad en el momento del diagnostico:

TABLA 2.3

Tiempo	q_t	p_t	$S_X(t)$
0	0.0000	1.0000	1.0000
2	0.0833	0.9167	0.9167
3	0.0909	0.9091	0.8333
6	0.2000	0.8000	0.6667
7	0.1250	0.8750	0.5833
10	0.1429	0.8571	0.5000
15	0.3333	0.6667	0.3333
16	0.2500	0.7500	0.2500
27	0.3333	0.6667	0.1667
30	0.5000	0.5000	0.0833
32	1.0000	0.0000	0.0000

La Gráfica 2.2 presenta la Curva de supervivencia mediante el método de Kaplan-Meier de los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes del hospital central de Guatemala de a lo sumo 40 años de edad en el momento del diagnostico.

GRÁFICA 2.2

Fuente: PREDOMINGO, Alejandro. Fundamentos Teóricos del Análisis de Sobrevivencia, Pág. 495.

Observación 2.3.1

Para incluir la información parcial sobre los tiempos censurados de supervivencia mediante el método de Kaplan - Meier, es posible que la curva de supervivencia sufra modificaciones.

Ejemplo 2.3

Supongamos que cuando se analizaron los datos del ejemplo 2.2, los pacientes con el segundo y sexto tiempo de supervivencia no habían muerto todavia. Se encontraban vivos después de 3 y 10 meses de seguimiento, respectivamente.

La tabla 2.4 muestra los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes de a lo sumo 40 años de edad en el momento del diagnostico , incluyendo las observaciones censuradas.

TABLA 2.4

Número del paciente	Supervivencia(meses)
1	2
2	3^+
3	6
4	6
5	7
6	10^+
7	15
8	15
9	16
10	27
11	30
12	32

Observe que los tiempos de supervivencia no cambian su valor anterior si la observación en el tiempo t esta censurada; no obstante, esta observación no se utiliza para calcular la probabilidad de un fallecimiento en cualquier punto del tiempo que sigue:

- Como el instante de tiempo 3 es censurado, se tiene:

$$q_3 = \frac{0}{11} = 0{,}0000$$
$$p_3 = \frac{12}{12} = 1{,}0000$$
$$S_X(3) = \frac{12-0}{12} * S_X(2) = 1{,}0000 * 0{,}9167 = 0{,}9167$$

- Para el instante de tiempo 10 que es censurado, se tiene:

$$q_{10} = \frac{0}{7} = 0{,}0000$$
$$p_{10} = \frac{7}{7} = 1{,}0000$$
$$S_X(10) = \frac{7-0}{7} * S_X(7) = 1{,}0000 * 0{,}6417 = 0{,}6417$$

TABLA 2.5
Método de Kaplan - Meier para calcular la función de supervivencia en los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes del hospital central de Guatemala de a lo sumo 40 años de edad en el momento del diagnostico, incluyendo las observaciones censuradas.

Tiempo	q_t	p_t	$S_X(t)$
0	0.0000	1.0000	1.0000
2	0.0833	0.9167	0.9167
3	0.0000	1.000	0.9167
6	0.2000	0.8000	0.7333
7	0.1250	0.8750	0.6417
10	0.0000	1.0000	0.6417
15	0.3333	0.6667	0.4278
16	0.2500	0.7500	0.3208
27	0.3333	0.6667	0.2139
30	0.5000	0.5000	0.1069
32	1.0000	0.0000	0.0000

En la gráfica 2.3 se presenta la curva de supervivencia mediante el método de Kaplan - Meier en los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 12 pacientes del hospital central de Guatemala de a lo sumo 40 años de edad en el momento del diagnostico, incluyendo las observaciones censuradas.

GRÁFICA 2.3

Fuente: PREDOMINGO, Alejandro. Fundamentos Teóricos del Análisis de Sobrevivencia, Pág. 496.

2.4. Método de Log-Rank.

Frecuentemente, se analizan los tiempos de supervivencia en una muestra de la población en estudio; mediante el método de Log-Rank o prueba del intervalo logarítmico, es posible comparar a través de tablas de contingencias, las distribuciones de los tiempos de supervivencia de dos o más muestras tomadas de una misma (o diferente) población, que posiblemente diferirán con respecto a cierto factor o variable cualitativa; aunque es posible la comparación de las distribuciones de los tiempos correspondientes a dos poblaciones diferentes.

Para llevar a cabo el método de Log-Rank, se utiliza el método de kaplan-Meier para obtener las funciones de supervivencia de cada muestra y las gráficas de las curvas de supervivencia; y posteriormente compararlas para detectar si existe alguna diferencia entre los tiempos de supervivencia. Si la diferencia se produce, las tasas de mortalidades entre los grupos varían consistentemente; es decir, la tasa de mortalidad de la primera muestra, es mayor a la tasa de mortalidad de la segunda muestra, y asi sucesivamente. Si no se presenta este caso o las curvas de supervivencia se cruzan, entonces esta prueba puede no ser válida para detectar diferencias cuando existan.
El objetivo de utilizar las tablas de contingencia se debe en primer lugar, que permiten registrar y analizar la relación entre los factores o variables cualitativas; y en segundo lugar, analizar si existe alguna relación de dependencia o independencia entre los niveles de las variables cualitativas objeto de estudio.

Definición 17. ***Tabla de Contingencia***.

La tabla de contingencia, es una tabla de doble entrada que se define por el número de factores o variables cualitativas que se analizan conjuntamente y el número de modalidades o niveles de los mismos. Consideremos, la siguiente tabla de contingencia de orden $h * k$; donde h es el número de factores o variables cualitativas $(A$ y $B)$ y k el número de niveles que contiene cada factor.

$A\|B$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_k	$n_{i.}$
A_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1k}	$n_{1.}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2k}	$n_{2.}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{ik}	$n_{i.}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_k	n_{k1}	n_{k2}	$\cdots$	n_{kj}	$\cdots$	n_{kk}	$n_{k.}$
$n_{.j}$	$n_{,1}$	$n_{,2}$	$\cdots$	$n_{.i}$	$\cdots$	$n_{.k}$	N

donde,

n_{ij} *: número de observaciones del estudio que presentaron (conjuntamente) el factor A en el i-ésimo instante de tiempo y el factor B en el j-ésimo instante de tiempo.*

$n_{i.}$ *: número de unidades del estudio que presentaron el factor A en el i-ésimo instante de tiempo (marginal i). Esto se obtiene de la siguiente manera:*

$$n_{i.} = \sum_{j=1}^{k} n_{ij} \tag{2.7}$$

$n_{.j}$ *: número de unidades del estudio que presentaron el factor B en el j-ésimo instante de tiempo (marginal j). Esto se obtiene de la siguiente manera:*

$$n_{.j} = \sum_{j=1}^{k} n_{ij} \tag{2.8}$$

N: número total de unidades observadas en el estudio. Esto es,

$$N = \sum_{i=1}^{h} n_{i.} = \sum_{j=1}^{k} n_{.j} = \sum_{i=1}^{h}\sum_{j=1}^{k} n_{ij} \tag{2.9}$$

A partir de una tabla de contingencia se puede además analizar si existe alguna relación de dependencia o independencia entre los niveles de los factores o variables cualitativas objeto de estudio.

Si se quisiera analizar conjuntamente la independencia de dos factores o variables cualitativas A y B, donde (A, B) se denomina atributo bidimensional de valores $(a_i,\ b_j)$, y junto con sus frecuencias absolutas[3], constituyen una distribución bidimensional. Se afirma que A y B son dos factores independientes, si se cumple:

a) Las frecuencias relativas condicionadas son iguales a las frecuencias relativas marginales. Para una tabla de contingencia $2 * k$,

   Las frecuencias relativas condicionadas, estas dadas de la siguiente manera:

$$f(A_1|B_1) = \frac{n_{11}}{n_{1.}} = \cdots = f(A_1|B_j) = \frac{n_{1j}}{n_{1.}} = \cdots = f(A_1|B_k) = \frac{n_{1k}}{n_{1.}} = \frac{n_{1.}}{N}$$

$$f(A_2|B_1) = \frac{n_{21}}{n_{2.}} = \cdots = f(A_2|B_j) = \frac{n_{2j}}{n_{2.}} = \cdots = f(A_2|B_k) = \frac{n_{2k}}{n_{2.}} = \frac{n_{2.}}{N}$$

   y así sucesivamente hasta obtener,

$$f(A_k|B_1) = \frac{n_{k1}}{n_{k.}} = \cdots = f(A_k|B_j) = \frac{n_{kj}}{n_{k.}} = \cdots = f(A_k|B_k) = \frac{n_{kk}}{n_{k.}} = \frac{n_{k.}}{N}$$

   Para las frecuencias relativas condicionadas $f(A_i|B_j)$ y $f(B_j|A_i)$, se tiene:

$$f(A_i|B_j) = \frac{n_{ij}}{n_{i.}} = f_{ij} = \frac{n_{i.}}{N}$$

$$f(B_j|A_i) = \frac{n_{ij}}{n_{.j}} = f_{ji} = \frac{n_{.j}}{N}$$

   donde, f_{ij} y f_{ij} son las frecuencias relativas marginales.

[3] La frecuencia absoluta para los factores A y B, puede interpretarse como el número de unidades observadas en cada factor respectivamente.

b) O bien, si se cumple que la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales:

$$f(A_i \cap B_j) = \frac{n_{ij}}{N} = \frac{n_{i.}}{N} * \frac{n_{.j}}{N}; \qquad \text{para} \quad \forall_{i,j} \tag{2.10}$$

donde, $f(A_i \cap B_j)$ es la frecuencia relativa conjunta de los factores A y B para todo instante de tiempo (i, j).

De la expresión (2.10), se puede afirmar que los valores del factor o variable cualitativa A, no están influidos por la modalidad o nivel que adopte el factor o variable cualitativa B.

De esta forma, comparando las frecuencias teóricas esperadas en caso de independencia entre las variables cualitativas con las frecuencias observadas en la muestra, se puede concluir que existe una relación de dependencia o independencia entre los factores o variables cualitativas analizadas.

Para identificar relaciones de dependencia entre los factores o variables cualitativas A y B, se utiliza un contraste de hipótesis estadístico, que según Ciro Martinez Bercardino, en su libro "Estadística y Muestreo" define que "una hipótesis estadística es una afirmación respecto a una característica de una población, y Contrastar una hipótesis es comparar las predicciones que se deducen de ella con la realidad que observamos: si hay coincidencia, dentro del margen de error admisible, mantendremos la hipótesis; en caso contrario, la rechazaremos. Rechazar una hipótesis implica sustituirla por otra capaz de explicar los datos observados". En el método de Log-Rank se utiliza el contraste de hipótesis estadístico o prueba de χ^2 (Chi cuadrado), cuyo cálculo nos permitirá afirmar con un nivel de confianza estadístico determinado si los niveles de un factor o variable cualitativa influyen en los niveles del otro factor o variable cualitativa analizada.

La Prueba de chi-Cuadrado $\hat{\chi}^2$, es la suma de fracciones que tienen por numerador el cuadrado de las diferencias entre las frecuencias reales u observadas y las frecuencias esperadas o teóricas y, por denominador la frecuencia esperada; esto es,

$$\hat{\chi}^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \tag{2.11}$$

donde,

e_{ij} : es el número de casos o frecuencia absoluta esperada en condiciones de independencia. Por lo tanto, las frecuencias esperadas se calculan de la siguiente manera:

$$e_{ij} = \frac{n_{i.} * n_{.j}}{N} \qquad \text{para} \quad \forall\, i, j. \tag{2.12}$$

Según la notación de la tabla de contingencia de orden $2 * k$ y utilizando el concepto frecuentalista de probabilidad, podemos estimar la probabilidad de que se de un suceso determinado, a partir de sus frecuencias relativas:

$$P(n_{ij}) = \frac{n_{ij}}{N}; \qquad P(n_{i.}) = \frac{n_{i.}}{N}; \qquad P(n_{.j}) = \frac{n_{.j}}{N}$$

donde,

$P(n_{ij})$: es la probabilidad de que conjuntamente los factores A y B hayan experimentado el evento de interés en los instantes de tiempo $i,\ j$, respectivamente.

De esta forma, si las variables son independientes,

$$P(\hat{n_{ij}}) = \frac{n_{i.}}{N} * \frac{n_{.j}}{N} = \frac{e_{ij}}{N}$$

Se puede observar en la expresión (2.11), que mientras mayor sea la coincidencia entre las frecuencias observadas n_{ij} y las esperadas e_{ij}, menor sera el valor de ji-cuadrado, lo que concluye que los factores o variables cualitativas son independientes, no pudiéndose afirmar lo mismo en caso contrario. Si $\chi^2 = 0$, significa que hay una completa concordancia entre las frecuencias observadas y las esperadas.

Nota:
Para la prueba de Chi-Cuadrado $\hat{\chi}^2$, aplicada a una tabla de contingencia de orden $k * j$, se tiene que los grados de libertad serán iguales a:

$$\nu = (k-1)(j-1)$$

Como se ejecuta el método de Log- Rank.

1. *Plantear la hipótesis nula y la hipótesis alternativa.*

   La hipótesis nula[4]a contrastar, denotada (H_0) será la de independencia entre los factores o variables cualitativas, es decir, el valor hipotético del parámetro que se compara con el resultado muy poco probable cuando la hipótesis es cierta, y siendo la hipótesis alternativa (H_1) la de dependencia entre los factores.

   En general, la hipotesis nula (H_0) se establece en términos de igualdad, es decir aquélla que postula que no existen diferencias en los tiempos de supervivencia de los grupos (dependencia entre los niveles de las variables cualitativas).

   Para establecer la hipótesis alternativa H_1, puede hacerse de tres (3) maneras, dependiendo del interés del investigador: en la primera se postula que son diferentes, la segunda una es mayor que la otra, y la tercera una es menor que la otra.

2. *Construir una tabla que contenga los valores observados de los grupos ordenados cronológicamente de menor a mayor para cada grupo.*

3. *Se grafican mediante el método de Kaplan-Meier las curvas de supervivencia para los grupos en estudio, y se comparan gráficamente.*

4. *Se construye la tabla de contingencia de orden $h * k$. Cuando se trabaja con tiempos de supervivencia, se obtiene una serie de tablas de contingencia que presentan el grupo en función del estado de supervivencia para cada tiempo t en el que sobreviene el evento de interés. Debajo de cada valor observado, se coloca el valor esperado utilizando la expresión (2.11).*

5. *Calcular el valor del estadístico de prueba $\hat{\chi}^2$, usando la expresión (2.12).*

6. *Determinar el valor crítico del estadístico de prueba $\hat{\chi}^2$, mediante la tabla[5] χ^2;* esto es:

   El valor de $\hat{\chi}^2$ calculado se compara con el valor tabulado de una χ^2 para un nivel de confianza α, determinado y ν grados de libertad. Si el valor calculado es mayor que el valor de tablas de una $\hat{\chi}^2_{(h-1)(k-1)}$, significará que las diferencias entre las frecuencias observadas (n_{ij}) y las frecuencias esperadas e_{ij}, son muy

[4]Cuando se considera una hipótesis nula (H_o), significa que el riesgo es el mismo en los grupos que se estan estudiando.

[5]Ver anexo B, tabla χ^2.

elevadas y por tanto se concluye que con un determinado nivel de confianza existe dependencia entre los factores o variables cualitativas analizadas. Esto es,

$\hat{\chi}^2 > \chi^2_{(h-1)(k-1)} \Rightarrow$ Rechazar hipótesis nula (dependencia entre los factores).
$\hat{\chi}^2 < \hat{\chi}^2_{(h-1)(k-1)} \Rightarrow$ Aceptar hipótesis nula (Independencia entre los factores).

Cuando la diferencia entre estos valores es cero 0, no hay asociación entre las variables cualitativas; pero además indica que son independientes.

Nota: *Para especificar el nivel de significancia que se va a utilizar, en necesario que al contrastar una cierta hipótesis, la máxima probabilidad con la que estamos dispuestos a correr el riesgo de cometer un error de tipo* 1 *(si rechazamos una hipótesis cuando debiera ser aceptada), se le llama nivel de significación. En la práctica, es frecuente un nivel de significación de* (0,05 , 0,01)*, si bien se unen otros valores. Si por ejemplo, se escoge el nivel de signigicación* 0,05 (5 %) *al diseñar una regla de decisión, entonces hay unas cinco* (5) *oportunidades entre* 100 *de rechazar la hipótesis cuando debiera haberse aceptado; es decir, tenemos* 95 % *de confianza de que hemos aceptado la decisión correcta. En tal caso decimos que la hipótesis ha sido rechazada al nivel de significación* $\alpha = 0{,}05$*, lo cual quiere decir que tal hipótesis tiene una probabilidad* 0,05 *de ser falsa.*

7. *Criterio de decisión*: si el valor crítico es menor que el valor del estadístico de prueba, rechazamos la hipótesis nula H_0. De lo contrario, aceptamos la hipótesis nula.

Ejemplo 2.4

Considérese los tiempos de supervivencia desde el diagnóstico de SIDA hasta la muerte de 21 pacientes del hospital central de Guatemala en el año 2005. Los datos se estratificaron por la edad de los pacientes en el momento del diagnóstico, y se obtuvieron los siguientes grupos:

- Grupo A: Los datos de 12 pacientes de a lo sumo 40 años de edad en el momento del diagnóstico. (Ver tabla 2.1).

- Grupo B: Los datos de 9 pacientes de más de 40 años de edad en el momento del diagnóstico.

TABLA 2.6

Edad > 40 años (Grupo B)	
Número del paciente	Supervivencia(meses)
1	1
2	1
3	1
4	1
5	2
6	3
7	3
8	9
9	22

Para el análisis de los tiempos de supervivencia de los grupos A y B, se utilizará el método de Log-Rang. Siguiendo el procedimiento, se tiene:

1. Se considera la hipótesis nula H_o aquella que postula que no existen diferencias en los tiempos de supervivencia de los grupos A y B; es decir,

$$H_o : S_{X,A}(t) = S_{X,B}(t)$$

   donde, $S_{X,A}$ y $S_{X,B}$ denotan la supervivencia para el grupo A y el grupo B, respectivamente.

   La hipótesis alternativa H_1, postula que existe alguna diferencia entre los tiempos de supervivencia de los grupos A y B; esto es:

$$H_1 : S_{X,A}(t) \neq S_{X,B}(t)$$

2. La tabla 2.7 contiene la unificación de las tablas 2.4. y 2.4.1.

   La tabla 2.7 muestra los tiempos de supervivencia (meses) desde el diagnóstico de SIDA hasta la muerte, de los 21 pacientes del hospital central de Guatemala en el año 2004, estratificados por edad (≤ 40 años, y > 40 años) en el momento del diagnóstico.

   TABLA 2.7

	Supervivencia (meses)	
No. del paciente	Edad $\leq$ 40 años	Edad $>$ 40 años
1	2	1
2	3	1
3	6	1
4	6	1
5	7	2
6	10	3
7	15	3
8	15	9
9	16	22
10	27	-
11	30	-
12	32	-

3. Grafiquemos y comparemos las curvas de supervivencia que producen los tiempos de supervivencia de cada grupo mediante el método de Kaplan - Meier:

   - Para el **grupo** A se tiene la tabla 2.3 y la gráfica 2.2 (realizadas anteriormente) que muestra la curva de supervivencia mediante el método de Kaplan- Meier para los 12 pacientes de a lo sumo 40 años de edad en el momento del diagnóstico.
   - Para gráficar la curva de supervivencia mediante el método de Kaplan-Meier para el **grupo** B, consideremos la siguiente tabla:

     TABLA 2.8
     Método de Kaplan - Meier para calcular la función de supervivencia en los tiempos de supervivencia (meses) de 9 pacientes con SIDA del hospital central de Guatemala en el año 2005, con más de 40 años de edad en el momento del diagnóstico.

Tiempo	q_t	p_t	$S_X(t)$
0	0.0000	1.0000	1.0000
1	0.4444	0.5556	0.5556
2	0.2000	0.8000	0.4445
3	0.5000	0.5000	0.2223
9	0.5000	0.5000	0.1112
22	1.0000	0.0000	0.0000

Teniendo en cuenta la tabla 2.8 se obtiene la curva de supervivencia mediante el método de Kaplan- Meier para los 9 pacientes de más de 40 años de edad en el momento del diagnóstico de SIDA, en el hospital central de Guatemala en el año 2005.

La Gráfica 2.4 muestra las Curvas de supervivencia de los dos grupos de pacientes estratificados por edad en el momento del diagnóstico en el hospital central de Guatemala en el año 2005.

GRÁFICA 2.4

Fuente: PREDOMINGO, Alejandro. Fundamentos Teóricos del Análisis de Sobrevivencia, Pág. 501.

Como puede verse en las curvas de supervivencia en cualquier punto en el tiempo, después del diagnóstico de SIDA, la probabilidad calculada de supervivencia es más alta en los pacientes que eran más jóvenes en el momento del diagnóstico.

4. Construir las tablas de contingencia de orden $2*2$, para comparar y analizar los tiempos de supervivencia en los que se observaron la muerte de los pacientes de los grupos A y B. ($t = 1, 2, 3, 6, 7, 9, 10, 15, 16, 22, 27, 30, 32$). Debajo de cada valor observado colocar el valor esperado.

    - Cuando $t = 1$ (mes), ningunó de los 12 pacientes del grupo A muere pero 4 del grupo B fallecen.

	DECESO		
Grupo	Si	No	Total
Edad ≤ 40	0 (2.29)	12	12
Edad > 40	4 (1.71)	5	9
Total	4	17	21

De la expresión (2.11), se tiene:

$$e_{11} = \frac{4 * 12}{21} = 2{,}29$$
$$e_{21} = \frac{4 * 9}{21} = 1{,}71$$

- Cuando $t = 2$ (mes), 1 de los 12 pacientes del grupo A muere, y 1 de los 5 pacientes restantes del grupo B también.

	DECESO		
Grupo	Si	No	Total
Edad ≤ 40	1 (1.41)	11	12
Edad > 40	1 (0.59)	4	5
Total	2	15	17

De la expresión (2.11), se tiene:

$$e_{11} = \frac{2 * 12}{17} = 1{,}41$$
$$e_{21} = \frac{2 * 5}{17} = 0{,}59$$

- Cuando $t = 3$ (mes), 1 de los 11 pacientes restantes del grupo A muere, y 2 de los 4 pacientes restantes del grupo B también.

	DECESO		
Grupo	Si	No	Total
Edad ≤ 40	1 (2.2)	10	11
Edad > 40	2 (0.8)	2	4
Total	3	12	15

De la expresión (2.11), se tiene:

$$e_{11} = \frac{3*11}{15} = 2{,}2$$
$$e_{21} = \frac{3*4}{15} = 0{,}8$$

y asi sucesivamente para todos los instantes de tiempo en los cuales se observaron la muerte de los 21 pacientes con SIDA del hospital central de Guatemala en el año 2005.

5. El valor de la estadística de prueba $\hat{\chi}^2$ está dada por:

$$\hat{\chi}^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = 15{,}98$$

Para obtener $\hat{\chi}^2$, la información obtenida se organiza mediante la siguiente tabla:

n_{ij}	e_{ij}	$(n_{ij}-e_{ij})^2$	$(n_{ij}-e_{ij})^2/e_{ij}$
0	2.29	-2.29	2.29
4	1.71	2.29	3.07
1	1.41	-0.41	0.12
1	0.59	0.41	0.28
1	2.2	-1.2	0.65
2	0.8	1.2	1.8
2	1.67	0.33	0.07
0	0.33	-0.33	0.33
1	0.8	0.2	0.05
0	0.2	-0.2	0.05
1	0.87	0.13	0.02
0	0.12	-0.12	0.12
0	0.78	-0.78	0.78
1	0.22	0.78	2.76
2	1.71	0.29	0.05

n_{ij}	e_{ij}	$(n_{ij}-e_{ij})^2$	$(n_{ij}-e_{ij})^2/e_{ij}$
0	0.29	-0.29	0.29
1	0.8	0.2	0.05
0	0.2	-0.2	0.2
0	0.75	-0.75	0.75
1	0.25	0.75	2.25
1	1	0	0
0	0	0	0
1	1	0	0
0	0	0	0
1	1	0	0
			15.98

6. Para determinar el valor crítico del estadístico de prueba χ^2, procedemos de la siguiente manera:

   - Determinar los grados de libertad:

   $$v = (h-1)(k-1) = (2-1)(2-1) = 1$$

   - Especificar el nivel de significancia que se va a utilizar:

   $$\alpha = 0{,}05$$

   El valor crítico del estadístico χ^2 para un nivel de confianza $\alpha = 0{,}05$ y grados de libertad $gl = 1$, denotado $\chi^2_{0,05}(1)$. En la tabla[6] ji - cuadrado encontramos que vale 3,84.

7. Criterio de desición:

   Como $\chi^2_{0,05}(1) = 3{,}84$ y el valor del estadístico de prueba $\hat{\chi}^2 = 15{,}98$; es decir que,

   $$\hat{\chi}^2 > \chi^2_{0,05}(1)$$

   Por lo tanto, rechazamos la hipótesis nula, es decir:

   $$H_0 : S_{X\leq 40}(t) = S_{X>40}(t)$$

   Concluimos, que la edad en la que se les fue diagnosticado el SIDA, es un factor determinante en que los individuos sobrevivan o no.

[6]Ver tabla ji cuadrado anexo B.

Conclusiones

1. Mediante una revisión a nivel introductorio de conceptos básicos del análisis de supervivencia como: la función de supervivencia, y la fuerza de mortalidad o de riesgo, a partir de la estimación e interpretación se puede realizar la descripción y resumen de los tiempos de supervivencia en el estudios de la supervivencia de una población. Se consideraron probabilidades condicionales de supervivencia, para estimar la probabilidad de un evento en un determinado intervalo de tiempo.

2. Mediante la descripción de las características propias de los métodos estadísticos no-paramétricos en el análisis de supervivencia, tales como el método de la tabla de vida, el método de Kaplar-Meier y el método de Log-Rank, se destacan ventajas y desventajas de la aplicación de estos métodos al análisis de tiempos de supervivencia de uno o varios grupos. Con el método de la tabla de vida se pueden analizar muestras relativamente pequeñas; puesto que al dividir los tiempos en intervalos de amplitud fija o variada se espera que al menos en cada intervalo haya evidencia del evento de interés; pero si por el contrario, la población es grande éste método no es el apropiado ya que puede suceder que en algunos intervalos no haya evidencia del evento de interés del estudio. Mientras, que una ventaja del método de Kaplan-Meier es que permite análizar cualquier tamaño de población, debido a que se estiman los tiempos en los que se produjo el evento de interés en la función de supervivencia. Por último, cuando el objetivo es comparar y analizar los tiempos de supervivencia de dos o más muestras de una misma (o diferente) población para el establecimiento y comprensión de independencia o dependencia entre los factores o variables cualitativas que puedan existir entre los tiempos de supervivencia de las muestras y las variables explicativas, el método de Log-Ragk es el más recomendado.

3. Con la elaboración de éste documento, los lectores interesados en el tema encontrarán una revisión de algunos métodos estadísticos en el análisis de supervivencia escrito de forma clara y simple, que podrá servir de base para posteriores estudios; además, contribuyo al crecimiento personal en conocimientos avanzados de la estadística.

Bibliografía

ARTIS, Manuel. *Estadística Actuarial de vida.* Universidad de Barcelona, Economia, 2005.

BATTEN, R.W. *Mortality. Table Construction.* Prentice Hall, New Yersey, 1.978.

BLANCO C., Liliana. *Probabilidad.* Ed. Universidad Nacional de Colombia, Bogotá 2004.

CASTELAZO A., Luis. Análisis de sobrevivencia pacientes de ginecología. Hospital de Ginecología, Reporte 4, México, 2001.

Centro de investigacin en salud poblacional, Instituto de salud pblica, Vol.42, No.3 Mayo - Junio de 2000, Mxico.

CERTUCHE V. Claudia Lorena. *Introducción al análisis de supervivencia*, Universidad del Cauca, 2004.

COLLET, D. *Modelling survival data in medical research.* London: Chapman and Hall, 1994.

EVANS, Michael J., ROSENTHAL, Jefrey S. *Probabilidad y Estadística.* La ciencia de la incertidumbre. Editorial Reverte, 2005.

FREEMAN, H.,*Finite Differences for Actuarial Students.* Cambridge University Press, 1.960.

HUERTAS C., Jaime Abel. *Cálculo Actuarial: Contingencias de vida individual.* Editorial Unibiblos, Primera Edición, 2001.

ICB digital. *Técnicas Estadísticas en el Análisis de Supervivencia.* Bogotá, 2005.

KLEINBAUM, David G. *Survival Analysis.* Springer. 1.984

LEE ET, Wang JW. *Satistical Methods for survival data analysis.* 3rd. ed. Belmont, CA: Lifetime learning publications, 2003.

MARTINEZ Bercardino, Ciro. *Estadística y Muestreo.* Editorial ecoe, octava edición, 1997.

Microsoft Student Encarta Premiun 2008.

MOLINERO LM. *Tiempo hasta que ocurre un suceso.* Análisis de Supervivencia, 2001. Disponible en: URL.

OBREGÓN S. Iván. *Teória de la probabilidad.* Actuaría, 2002.

PEREZ, Cesar. *Técnicas Estadísticas con SPSS*, 2004.

PREDOMINGO, Alejandro. LETON, Emilio. *Fundamentos Teóricos del Análisis de Supervivencia.* Glaxowellcome, 1977.

RESTREPO S., Carlos Julio. *Teoría de la Probabilidad y Aplicaciones.* Editorial Universidad del Cauca. Primera Edición, 2006.

TURLER, H., *Matemática del Seguro.* Editorial Siglo XX, Bogotá, 1.972.

http://www.econ.au.dk/ecbologna/papers/IbsCI.pdf.

http://es.wikipedia.org/w/index.php.

htpp://webdelprofesor.ula.ve.economia.oliverosm. /materiasdictadas.produccion2. tabladistrnorm.pdf.

ANEXOS

ANEXO A

Tabla Colombiana de Mortalidad de los asegurados durante 84 - 89

x	l_x	d_x	q_x	p_x	e_x	$D_x(i)$	$C_x(i)$	$M_x(i)$	$R_x(i)$
20	100.000	345	0.003450	0.996550	52.2	14.864.3628	46.6200	597.8363	9.221.9564
21	99.655	345	0.003462	0.996538	51.4	13.466.4370	42.3819	551.2163	8.624.1200
22	99.310	346	0.003484	0.996516	50.5	12.199.8336	38.6406	508.8344	8.072.9038
23	98.964	347	0.003506	0.996494	49.7	11.052.1172	35.2294	470.1938	7.564.0694
24	98.617	348	0.003529	0.996471	48.9	10.012.1499	32.1190	434.9644	7.093.8756
25	98.269	348	0.003541	0.996459	48.1	9.069.8354	29.1991	402.8454	6.658.9112
26	97.921	349	0.003564	0.996436	47.2	8.216.1059	26.6209	373.6463	6.256.0658
27	97.572	350	0.003587	0.996413	46.4	7.442.5662	24.2702	347.0254	5.882.4195
28	97.222	350	0.003600	0.996400	45.6	6.741.6991	22.0638	322.7552	5.535.3942
29	96.872	351	0.003623	0.996377	44.7	6.106.7536	20.1153	300.6914	5.212.6390
30	96.521	352	0.003647	0.996353	43.9	5.531.4789	18.3387	280.5761	4.911.9476
31	96.169	354	0.003681	0.996319	43.0	5.010.2784	16.7663	262.2374	4.631.3715
32	95.815	355	0.003705	0.996295	42.2	4.538.0323	15.2852	245.4711	4.369.1341
33	95.460	357	0.003740	0.996260	41.4	4.110.1987	13.9739	230.1859	4.123.6630
34	95.103	358	0.003764	0.996236	40.5	3.722.5704	12.7391	216.2120	3.893.4771
35	94.745	358	0.003779	0.996221	39.7	3.371.4158	11.5810	203.4729	3.677.2651
36	94.387	359	0.003803	0.996197	38.8	3.053.3425	10.5576	191.8919	3.473.7922
37	94.028	361	0.003839	0.996161	38.0	2.765.2083	9.6513	181.3343	3.281.9003
38	93.667	364	0.003886	0.996114	37.1	2.504.1744	8.8468	171.6830	3.100.5659
39	93.303	367	0.003933	0.996067	36.2	2.267.6754	8.1088	162.8362	2.928.8829
40	92.936	371	0.003992	0.996008	35.4	2.053.4142	7.4520	154.7274	2.766.0467
41	92.565	376	0.004062	0.995938	34.5	1.859.2882	6.8659	147.2754	2.611.3193
42	92.189	383	0.004155	0.995845	33.7	1.683.3961	6.3579	140.4095	2.464.0439
43	91.806	394	0.004292	0.995708	32.8	1.524.0022	5.9459	134.0516	2.323.6344
44	91.412	408	0.004463	0.995537	31.9	1.379.5107	5.5974	158.1057	2.189.5828

x	l_x	d_x	q_x	p_x	e_x	$D_x(i)$	$C_x(i)$	$M_x(i)$	$R_x(i)$
45	91.004	427	0.004692	0.995308	31.1	1.248.5032	5.3255	122.5083	2.061.4771
46	90.577	452	0.004990	0.995010	30.2	1.129.6773	5.1249	117.1827	1.938.9688
47	90.125	485	0.005381	0.994619	29.4	1.021.8545	4.9991	112.0579	1.821.7861
48	89.640	525	0.005857	0.994143	28.5	923.9596	4.9991	112.0579	1.821.7861
49	89.115	570	0.006396	0.993604	27.7	835.0438	4.8556	102.1393	1.602.6695
50	88.545	619	0.006991	0.993009	26.9	754.2751	4.7936	97.2837	1.500.5302
51	87.926	669	0.007609	0.992391	26.1	680.9110	4.7098	92.4901	1.403.2465
52	87.257	718	0.008229	0.991771	25.3	614.3002	4.5953	87.7802	1.310.7564
53	86.539	763	0.008817	0.991183	24.5	553.8594	4.4394	83.1850	1.222.9762
54	85.776	803	0.009362	0.990638	23.7	499.0692	4.2473	78.7456	1.139.7912
55	84.973	835	0.009827	0.990173	22.9	499.4519	4.0151	74.4983	1.061.0456
56	84.138	858	0.010198	0.989802	22.1	404.5776	3.7506	70.4832	986.5474
57	83.280	873	0.010483	0.989517	21.3	364.0472	3.4693	66.7325	916.0642
58	82.407	887	0.010764	0.989236	20.6	327.4827	3.2045	63.2633	849.3317
59	81.520	910	0.011163	0.988837	19.8	294.5071	2.9887	60.0588	789.0684
60	80.610	951	0.011798	0.988202	19.0	264.7450	2.8394	57.0701	726.0096
61	79.659	1.012	0.012704	0.987296	18.2	237.8379	2.7468	54.2307	668.9395
62	78.647	1.090	0.013859	0.986141	17.4	213.4694	2.6896	51.4839	614.7088
63	77.557	1.184	0.015266	0.984734	16.7	191.3735	2.6559	48.7943	563.2249
64	76.373	1.292	0.016917	0.983083	15.9	171.3200	2.6347	46.1383	514.4307
65	75.081	1.408	0.018753	0.981247	15.2	153.1107	2.6103	43.5036	468.2923
66	73.673	1.532	0.020795	0.979205	14.5	136.5813	2.5320	40.8933	424.7888
67	72.141	1.655	0.022941	0.977059	13.8	121.5828	2.5357	38.3114	383.8955
68	70.486	1.769	0.025097	0.974903	13.1	107.9942	2.4640	35.7757	345.5841
69	68.717	1.874	0.027271	0.972729	12.4	95.7126	2.3729	33.3117	309.8084
70	66.843	1.973	0.029517	0.970483	11.7	84.6385	2.2712	30.9388	276.4967
71	64.870	2.076	0.032002	0.967998	11.1	74.6729	2.1725	28.6676	245.5579
72	62.794	2.182	0.034749	0.965251	10.4	65.7120	2.0758	26.4952	216.8903
73	60.612	2.294	0.037847	0.962153	9.8	57.6624	1.9840	24.4194	190.3951
74	58.318	2.457	0.042131	0.957869	9.2	50.4364	1.9318	22.4354	165.9758
75	55.861	2.636	0.047189	0.952811	8.5	43.9195	1.8841	20.5036	143.5404
76	53.225	2.888	0.054260	0.945740	7.9	38.0427	1.8766	18.6195	123.0367
77	50.337	2.951	0.058625	0.941375	7.4	32.7077	1.7432	16.7430	104.4172
78	47.386	3.107	0.065568	0.934432	6.8	27.9911	1.6685	14.9998	87.6742
79	44.279	3.365	0.075995	0.924005	6.2	23.7780	1.6427	13.3313	72.6744
80	40.914	3.514	0.085887	0.914113	5.7	19.9736	1.5595	11.6886	59.3430

x	l_x	d_x	q_x	p_x	e_x	$D_x(i)$	$C_x(i)$	$M_x(i)$	$R_x(i)$
81	37.400	3.685	0.098529	0.901471	5.2	16.5983	1.4867	10.1291	47.6544
82	33.715	3.878	0.115023	0.884977	4.7	13.6026	1.4224	8.6423	37.5254
83	29.837	4.054	0.135872	0.864128	4.3	10.9437	1.3518	7.2199	28.8831
84	25.783	4.075	0.158050	0.841950	3.8	8.5970	1.2352	5.8682	21.6631
85	21.708	4.059	0.186982	0.813018	3.5	6.5802	1.1185	4.6330	15.7949
86	17.649	3.692	0.209190	0.790810	3.2	4.8635	0.9249	3.5144	11.1620
87	13.957	3.268	0.234148	0.765852	2.9	3.4965	0.7443	2.5895	7.6475
88	10.689	2.803	0.262232	0.737768	2.6	2.4343	0.5803	1.8453	5.0580
89	7.886	2.315	0.293558	0.706442	2.3	1.6327	0.4357	1.2649	3.2128
90	5.571	1.832	0.328846	0.671154	2.1	1.0486	0.3135	0.8292	1.9479
91	3.739	1.377	0.368280	0.631720	1.9	0.6398	0.2142	0.5157	1.1187
92	2.362	974	0.42362	0.587638	1.7	0.3674	0.1377	0.3015	0.6029
93	1.388	641	0.461816	0.538184	1.5	0.1963	0.0824	0.1638	0.3014
94	747	386	0.516734	0.483266	1.3	0.0960	0.0451	0.0814	0.1376
95	361	209	0.578947	0.421053	1.1	0.0422	0.0222	0.0363	0.0562
96	152	99	0.651316	0.348684	1.0	0.0161	0.0096	0.0141	0.0199
97	53	39	0.735849	0.264151	0.8	0.0051	0.0034	0.0045	0.0058
98	14	12	0.857143	0.142857	0.6	0.0012	0.0010	0.0011	0.0012
99	2	2	1.000000	0.000000	0.5	0.0002	0.0001	0.0001	0.0001

Fuente: HUERTAS, J. Abel, contingencias de vida individual: Calculo Actuarial. Editorial Universidad Nacional, Medellin, 2005.

ANEXO B

Tabla de la Distribución Chi Cuadrado χ^2.
P la probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, v = Grados de Libertad.

$v\|p$	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152
6	22,4575	20,2491	18,5475	16.8119	14,4494	12,5916	10,6446	9,4461
7	24,3213	22,0402	20,2777	18,4753	16,4753	14,0671	12,0170	10,7479
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339

Fuente: htpp:webdelprofesor.ula.ve.economia.oliverosm.
materiasdictadas.produccion2.tabladistrnorm.pdf