

UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO

PROGRAMA DE MAESTRÍA

MAESTRÍA EN INFORMÁTICA



TESIS

**MODELO DE WEIBULL CON CENSURA INTERVALAR PARA EL
ANÁLISIS DE SOBREVIVENCIA DE PERSONAS CON CÁNCER DE
MAMA DE LA CLÍNICA AMERICANA DE JULIACA, 2016**

PRESENTADA POR

ELIZABETH MILAGROS GORDILLO MONTALVO

PARA OPTAR EL GRADO ACADÉMICO DE

MAGISTER SCIENTIAE EN INFORMÁTICA

MENCIÓN EN MATEMÁTICA Y SIMULACIÓN COMPUTACIONAL

PUNO, PERÚ

2017

UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO
PROGRAMA DE MAESTRÍA
MAESTRÍA EN INFORMÁTICA



TESIS

MODELO DE WEIBULL CON CENSURA INTERVALAR PARA EL
ANÁLISIS DE SOBREVIVENCIA DE PERSONAS CON CÁNCER DE
MAMA DE LA CLÍNICA AMERICANA DE JULIACA, 2016

PRESENTADA POR:

ELIZABETH MILAGROS GORDILLO MONTALVO

PARA OPTAR EL GRADO ACADÉMICO DE:
MAGISTER SCIENTIAE EN INFORMÁTICA
MENCIÓN EN MATEMÁTICA Y SIMULACIÓN COMPUTACIONAL

APROBADA POR EL SIGUIENTE JURADO:

PRESIDENTE

Dr. EDGAR ELOY CARPIO VARGAS

PRIMER MIEMBRO

M.Sc. SAMUEL DONATO PEREZ QUISPE

SEGUNDO MIEMBRO

M.Sc. LUIS HUGO HUACASI VASQUEZ

ASESOR DE TESIS

Dr. ALEJANDRO APAZA TARQUI

Puno, 14 de julio del 2017.

ÁREA: Modelación matemática.

TEMA: Modelación algorítmica de sistemas dinámicos.

DEDICATORIA

En primer lugar, quiero agradecer a Dios, por todas sus bendiciones, salud, fuerza y perseverancia para que pudiese concluir con este trabajo. A mi querido esposo José Luis y mis padres: Alejandro y Faustina, a mis hermanos: Julián y Gabriela, mi maravillosa familia, a quienes debo demasiado como para poder expresarlo con palabras, que han estado a mi lado en todo momento, empujándome a mejorar en mi vida, con su inmenso cariño.



AGRADECIMIENTOS

- A la Universidad Nacional Del Altiplano por poner el conocimiento a nuestro alcance y proporcionarnos el espacio para desarrollarnos como personas y forjadores del mañana.
- A nuestros docentes cuyas experiencias y saberes han transmitido desinteresadamente.



ÍNDICE GENERAL

	Pág.
DEDICATORIA.....	i
AGRADECIMIENTOS.....	ii
ÍNDICE GENERAL.....	iii
ÍNDICE DE CUADROS.....	v
ÍNDICE DE FIGURAS.....	vi
ÍNDICE DE ANEXOS.....	vii
RESUMEN.....	viii
ABSTRACT.....	ix
INTRODUCCIÓN.....	x
CAPÍTULO I	
PROBLEMÁTICA DE INVESTIGACIÓN	
1.1 PLANTEAMIENTO DEL PROBLEMA.....	1
1.2 JUSTIFICACIÓN.....	2
1.3 OBJETIVOS.....	3
1.3.1 Objetivo general.....	3
1.3.2 Objetivos específicos.....	3
1.4 HIPÓTESIS.....	3
CAPÍTULO II	
MARCO TEÓRICO	
2.1 FUNCIÓN DE SOBREVIVENCIA.....	5
2.2 CENSURA.....	8
2.3 DISTRIBUCIÓN DE WEIBULL.....	12
2.4 ANTECEDENTES.....	14

CAPÍTULO III**METODOLOGÍA**

3.1	ÁMBITO O LUGAR DE ESTUDIO	18
3.2	.POBLACIÓN Y MUESTRA	18
3.3	METODOLOGÍA	18

CAPÍTULO IV**RESULTADOS Y DISCUSIÓN**

4.1	ESTIMACIÓN DE LA FUNCIÓN DE SOBREVIVENCIA	35
4.2	COMPARACIÓN DE FUNCIONES DE SOBREVIVENCIA	37
4.3	AJUSTE DEL MODELO DE WEIBULL	39
4.4	VERIFICACIÓN DEL MODELO DE WEIBULL	42
4.5	RESIDUO	43
4.5.1	Comprobación de la hipótesis global del modelo: residuos de padronizados	43
4.5.2	Comprobación de la influencia sobre cada observación en el modelo: residuos de deviance	45
	CONCLUSIONES	46
	RECOMENDACIONES	48
	BIBLIOGRAFÍA	49
	ANEXOS	52

ÍNDICE DE CUADROS

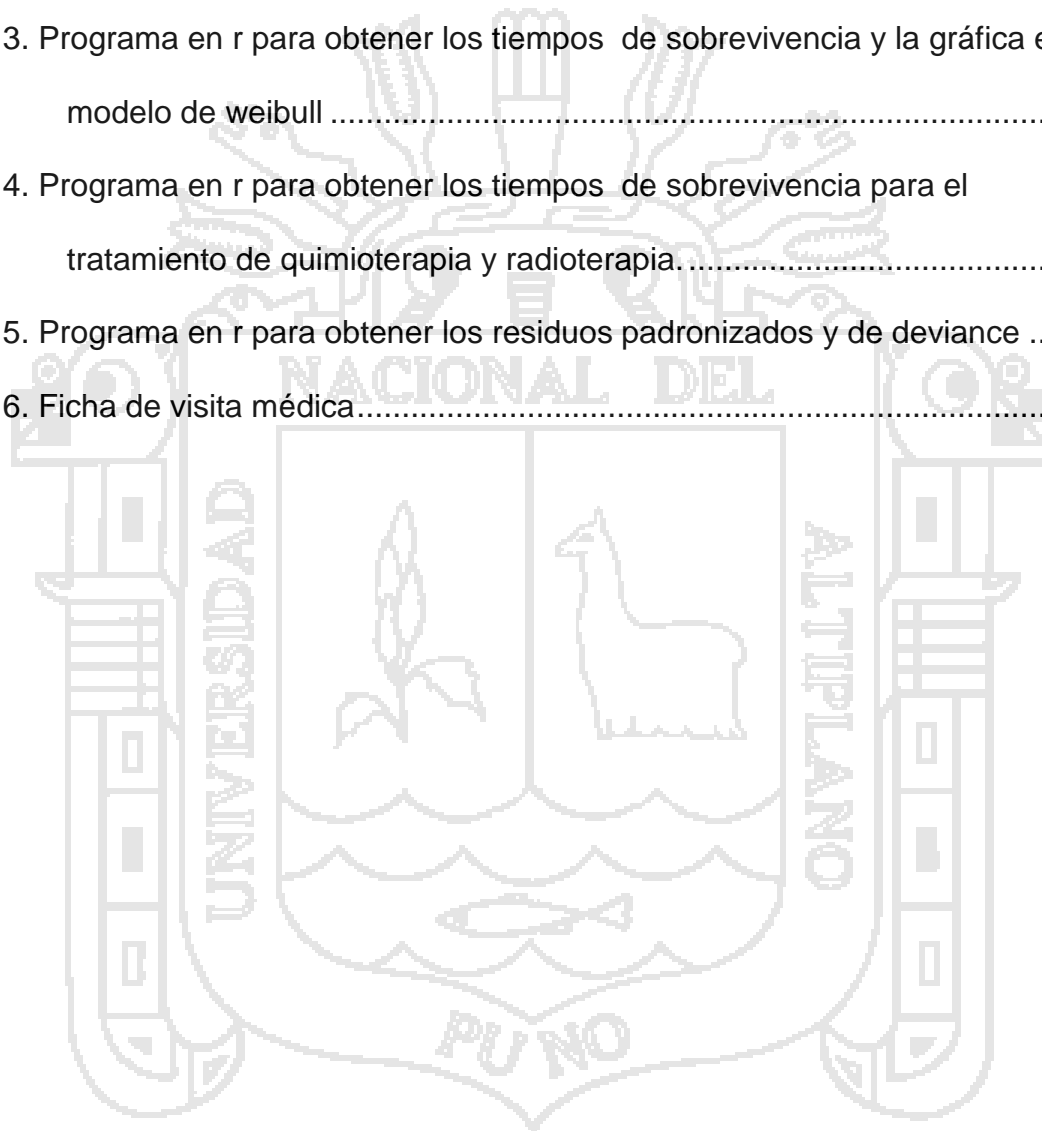
	Pág.
1. Tabla de contingencia para el contraste de igualdad de funciones de supervivencia en dos grupos en el instante t_i , donde los grupos 1 y 0 indican cada una de las funciones.	21
2. Información sobre el conjunto de datos de cáncer de mama. se incluye el valor mínimo, máximo, media, desviación típica y mediana de la edad de los pacientes en estudio.	34
3. Tiempos y probabilidad de supervivencia con el modelo de weibull.	36
4. Resultado de contraste de igualdad de funciones de supervivencia para la covariable tratamiento.	39
5. Parámetros del modelo de weibull para el estudio.	40
6 Resultado de contraste del modelo de weibull para cada covariable se incluye el p-value resultante del contraste.	43

ÍNDICE DE FIGURAS

	Pág.
1. Monotonía de la función de riesgo	8
2. Varios tiempos de censura.....	10
3. Función de sobrevivencia de la distribución weibull.....	14
4. Ejemplo de grafica de residuos de cox-snell, en la cual se representa el riesgo acumulado hr frente a los residuos r.....	26
5. Ejemplo de grafica de residuos de deviance, en la cual se representan los residuos frente al índice de cada paciente.....	27
6. Ejemplo de grafica de residuos de deviance frente a la puntuación de riesgo β^x para cada paciente.....	28
7. Estimación de weibull de la función de sobrevivencia.....	37
8. Estimación de weibull de la función de sobrevivencia para pacientes que recibieron tratamiento de quimioterapia y radioterapia.....	38
9. Curvas de sobrevivencia estimadas por el modelo de weibull versus la curva de sobrevivencia por kaplan-meier	42
10 Residuos de cox-snell para datos de cáncer de mama en el estudio de sobrevivencia.....	44
11. Residuos de deviance para datos de cáncer de mama en el estudio de sobrevivencia para determinar la existencia de outliers en el modelo.	45

ÍNDICE DE ANEXOS

	Pág.
1. Datos referentes al el tiempo de vida para pacientes con cáncer de mama.	53
2. Programa en r para obtener los parámetros del modelo de weibull	56
3. Programa en r para obtener los tiempos de sobrevivencia y la gráfica en el modelo de weibull	57
4. Programa en r para obtener los tiempos de sobrevivencia para el tratamiento de quimioterapia y radioterapia.....	59
5. Programa en r para obtener los residuos padronizados y de deviance	60
6. Ficha de visita médica.....	62



RESUMEN

El análisis de sobrevivencia trata de la evaluación estadística de variables que miden el tiempo hasta un evento de interés. En el área de estudios clínicos, este evento es muchas veces el inicio de una enfermedad, la desaparición de los síntomas o la muerte. Una particularidad, la cual ha de considerarse en el análisis de sobrevivencia, son datos censurados. Estos aparecen cuando el tiempo de interés no puede ser observado exactamente y la información al respecto es parcial. El presente proyecto se realizó en la Clínica Americana durante el año 2016, teniendo en cuenta el siguiente objetivo de determinar el modelo probabilístico de sobrevivencia con censura intervalar de los pacientes con cáncer de mama mediante el modelo de Weibull, para lo cual se determinó los parámetros de forma y de escala mediante los estimadores de máxima verosimilitud, y los tiempos con censura intervalar, para lo cual se utilizó como muestra las pacientes con cáncer de mama de la Clínica Americana. A partir de los resultados obtenidos, se determinó que la probabilidad de sobrevivencia está relacionada con el tipo de tratamiento, en donde se demuestra que la quimioterapia tiene una probabilidad de 96,77% y la radioterapia tiene una probabilidad de 90,48%, Se encontró que el parámetro de escala es 10.21737 y de forma es 2.73. La probabilidad de sobrevivencia para el tiempo de censura intervalar de un mes es de 97,37%, y para 16 meses es de 0,026%, esto es para pacientes con cáncer de mama de grado 3 y 4.

Palabras clave

Cáncer de mama, censura intervalar, modelo de Weibull, sobrevivencia, tiempo censurado.

ABSTRACT

Survival analysis is the statistical evaluation of variables that measure the time until an event of interest. In the clinical trials area, this event is often the illness beginning, the disappearance of the symptoms or death. A special feature, which has to be considered in the survival analysis, these are censored data. They appear when the interest time cannot be realized exactly and the information is partial. This project was conducted in the Americana clinic during 2016, considering the following objective of determining the survival probabilistic model with censorship intervalar of cancer breast patients using the Weibull model, for the which determined the parameters of form and scale using the maximum likelihood estimators, and intervalar censorship times, which was used as sample breast cancer patients of Americana clinic. From the results obtained, it was determined that the survival probability is related to the type of treatment, where is shown that chemotherapy has a 96,77% chance and the radiotherapy has a probability of 90,48%, It was found that the scale parameter is 10.21737 and the form is 2.73. The probability of survival time of one month intervalar censorship is 97,37% for 16 months is 0.026%, this is for breast cancer patients with grade 3 and 4.

Keywords

Breast cancer, censored time, interval censorship, survival, Weibull model.

INTRODUCCIÓN

El análisis de sobrevivencia es una expresión utilizada para designar un análisis estadístico de datos cuando la variable es estudio representa el tiempo desde un instante inicial, bien definido, hasta que ocurre un determinado evento de interés. Los conjuntos de datos de sobrevivencia son caracterizados por los tiempos de vida o tiempos hasta que ocurra el evento de interés y muy frecuentemente por las censuras.

En las enfermedades crónicas, tales como el cáncer, la sobrevivencia se mide como una probabilidad de permanecer vivo durante una determinada cantidad de tiempo. La sobrevivencia al año o a los 5 años es a menudo expresada como indicadores de la severidad de una enfermedad y como pronóstico. Típicamente, el pronóstico del cáncer se valora determinando el porcentaje de pacientes que sobrevive al menos cinco años después del diagnóstico.

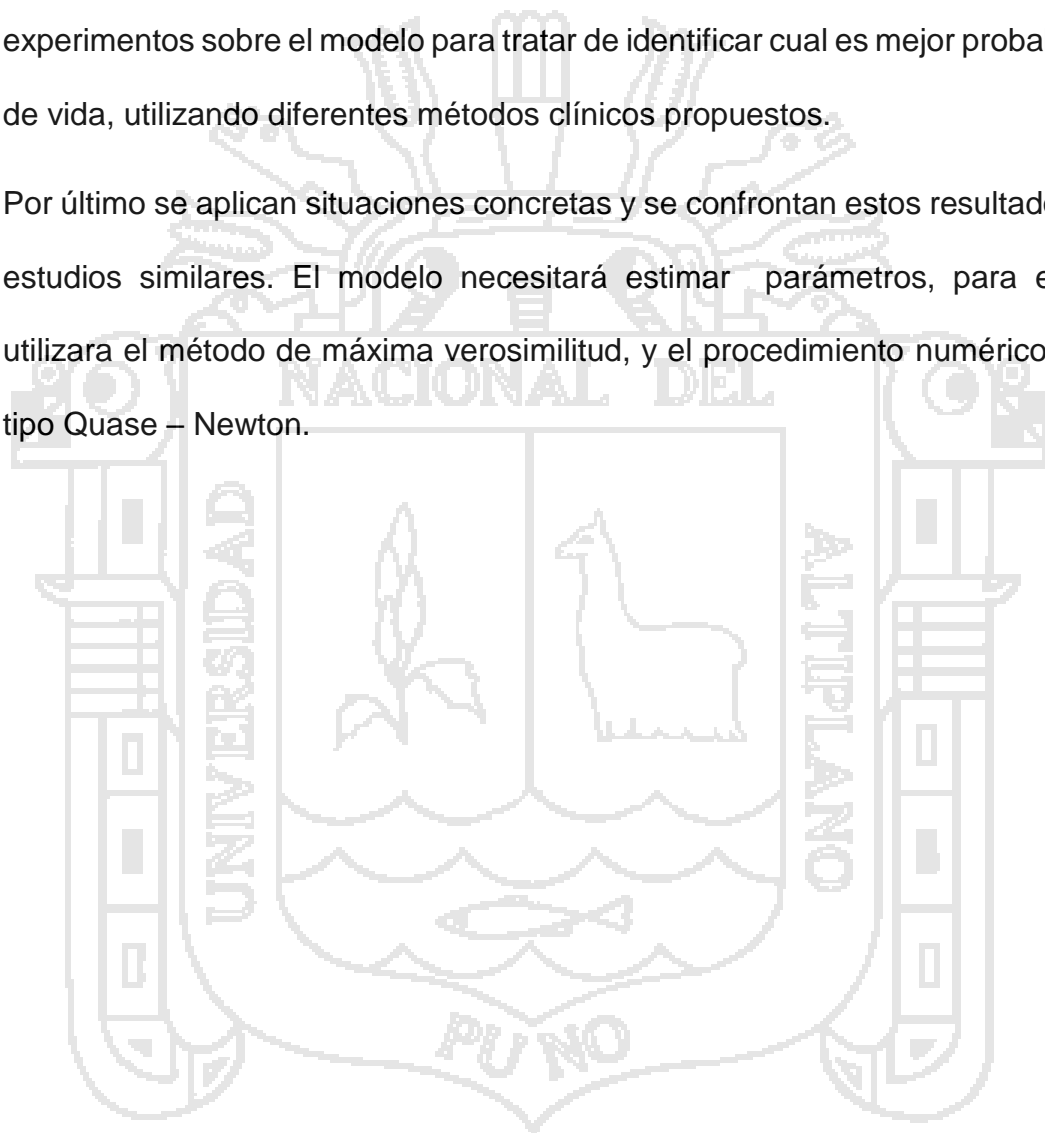
Como mencionamos anteriormente, uno de los problemas del análisis de tiempo de vida es la posibilidad de que estén censurados. La censura ocurre cuando no es posible observar el evento de interés durante el periodo de tiempo en que el individuo está en observación. La censura puede suceder por los siguientes motivos: la muerte del individuo sucede fuera del periodo de la duración del estudio o no fue posible observar el evento de interés.

En este trabajo presentamos el modelo Weibull para datos con censura intervalar construyendo la función de verosimilitud de la misma forma que en (Biswarbrata, 2012). La ventaja de esta distribución es que posee casos particulares de otras distribuciones usuales en el análisis de sobrevivencia, otra característica es la flexibilidad -de la función de tasa de falla para modelar diferentes formas, tales como, creciente, decreciente, unimodal y forma de bañera.

En la primera parte se identifica la problemática y la necesidad de investigar las posibilidades que se abren, y se plantea un modelo el cual permitirá encontrar la probabilidad de sobrevivencia para datos con censura intervalar de manera muy sencilla.

En la segunda parte se propone los métodos para describir los resultados de los experimentos sobre el modelo para tratar de identificar cual es mejor probabilidad de vida, utilizando diferentes métodos clínicos propuestos.

Por último se aplican situaciones concretas y se confrontan estos resultados con estudios similares. El modelo necesitará estimar parámetros, para ello se utilizará el método de máxima verosimilitud, y el procedimiento numérico es de tipo Quase – Newton.



CAPÍTULO I

PROBLEMÁTICA DE INVESTIGACIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

Tiempo de sobrevivencia es un tema principal en estadística médica y muchas razones hacen que sea difícil obtener información en estudios de tiempo de sobrevivencia. Un estudio a menudo se acaba antes de la muerte de todos los pacientes, y podremos mantener sólo la información que algunos pacientes estaban todavía vivos al final del estudio, sin tener en cuenta cuando realmente murieron. Eso es una motivación de la teoría del estudio de datos censurados. Vamos a ver que una primera posibilidad para hacer frente a no completar (vamos a decir censurado) datos debe ser consciente de ellos y calcular la estadística sólo en el resto de los datos. Sin embargo podemos perder alguna información ignorando datos censurados, y así nuestra estimación sería parcial porque ignorando datos censurados por la derecha por ejemplo está ignorando datos que tiene la propiedad de ser mayor que un valor dado.

Debido a la falta de sistemas clínicos de control para la sobrevivencia de pacientes con cáncer de mama y a la presencia de observaciones incompletas o parciales (censura), consideraremos el modelo de Weibull, cuya característica es la flexibilidad de su función de riesgo, la cual asume diferentes formas, tales como creciente, decreciente, unimodal. La investigación pretende responder a las siguientes preguntas: ¿Cómo es el modelo de Weibull con censura intervalar para el análisis de sobrevivencia de personas con cáncer de mama de la Clínica Americana-Juliaca 2016?, ¿Cómo ocurre el evento de interés, para las personas que están siendo observadas?, ¿Cuáles son los valores de los parámetros de escala y forma del modelo de Weibull?

1.2 JUSTIFICACIÓN

Cada año aproximadamente 1,3 millones de mujeres sufren de cáncer de mama en todo el mundo, lo que es la neoplasia más frecuente en las mujeres. Es por ello que nos lleva a estudiar un análisis de sobrevivencia, y así se pueda aplicar el método adecuado para tener menos muertes.

Debido a que la variable tiempo es una variable continua podría ser, en principio, estudiada mediante las técnicas de análisis de la varianza o los modelos de regresión. Hay, sin embargo, dos dificultades importantes para este planteamiento. En primer lugar, en la mayor parte de los estudios citados la variable tiempo no tiene una distribución normal, más bien suele tener una distribución asimétrica y aunque podrían intentarse transformaciones que la normalizaran, existe una segunda dificultad que justifica un planteamiento específico para estas variables, y es que para observarlas se tiene que prolongar el estudio durante un período de tiempo

suficientemente largo, en el cual suelen ocurrir pérdidas, que imposibilitan la observación del evento.

El presente proyecto tiene como finalidad mejorar los estudios clínicos y llevar un mejor control de sobrevivencia en los pacientes con cáncer de mama, también analizaremos el comportamiento de la función de riesgo. Para modelar este tipo de datos existen muchos modelos ya sean paramétricos, semi - paramétricos, para lo cual usaremos el modelo de Weibull que se ajustara a nuestros datos.

1.3 OBJETIVOS

1.3.1 Objetivo general

Determinar el modelo probabilístico de sobrevivencia con censura intervalar de los pacientes con cáncer de mama mediante el modelo de Weibull de la Clínica Americana - Juliaca 2016.

1.3.2 Objetivos específicos

- Determinar los parámetros de escala y de forma de la función de sobrevivencia del modelo de Weibull utilizando el paquete R.
- Determinar los tiempos con censura intervalar para los pacientes con cáncer de mama.

1.4 HIPÓTESIS

El modelo de Weibull con censura intervalar de datos permite determinar la probabilidad de sobrevivencia para personas con cáncer de mama.

CAPÍTULO II

MARCO TEÓRICO

El análisis de sobrevivencia es un -procedimiento estadístico implementado con mayor frecuencia a mediados del siglo XX. Este procedimiento estadístico es muy utilizado en la medicina, pero también se puede aplicar a otros campos como la ingeniería, sociología, psicología, educación, etc... Puede tener como objeto de estudio, el tiempo hasta que un automóvil tuvo su primera avería mecánica después de su compra. (Colosimo, 2008)

El análisis de sobrevivencia consiste, entre otras cosas, en analizar los tiempos de vida de los individuos desde el momento en que entro en estudio, hasta el instante en que ocurre el evento de interés, este evento es generalmente definido como una falla, que podría ser muerte, recaída de una enfermedad, o también cuando el tratamiento comienza hacer efecto en el paciente.

Lo que distingue al análisis de sobrevivencia de otras áreas de la estadística es la presencia de censura la cual es una característica de los datos de tiempo de vida.

Por censura entendemos que es el tiempo de falla de un individuo, es una información parcial sobre el tiempo de falla, existen varios tipos de censura, esto es, censura derecha, izquierda, intervalar.

Otro aspecto importante es tener en cuenta, el uso de las covariables, el tiempo de vida de los individuos es afectado por estas variables.

2.1 FUNCIÓN DE SOBREVIVENCIA

Comenzamos por definir el tiempo vida de un determinado individuo, de una población homogénea, que será representado por una variable aleatoria T , no negativa y absolutamente continua. (Colosimo, 2008). Podemos entonces definir la función de sobrevivencia como la probabilidad de sobrevivir para un instante t y será representado por:

$$S(t) = P(T > t), t \geq 0 \quad (1)$$

que es una función monótona, no creciente, continua y que tiene las siguientes propiedades:

1. $S(0)=1$,
2. $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$

Esto es, la probabilidad de sobrevivir al tiempo cero es uno y la de sobrevivir un tiempo infinito es nula.

La función densidad de probabilidad en un instante t y la tasa de instantánea de muerte en ese instante:

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt)}{dt} \tag{2}$$

Una función igualmente utilizada en este ámbito, es la función de riesgo, también conocida como la función tasa de falla o mortalidad, se describe como la tasa instantánea de muerte de un individuo, que sobrevive hasta el instante t, dada por:

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

Aplicando la ley de probabilidad condicional a la ecuación (2), se obtiene

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

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

Dado que

$$P(t \leq T < t + dt) = \int_t^{t+dt} f(u) du = F(t + dt) - F(t)$$

Al sustituir (1) y (2) aplicando la definición de derivada, se obtiene

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

$$h(t) = \lim_{dt \rightarrow 0} \frac{F(t + dt) - F(t)}{P(T \geq t) dt} = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)}$$

De manera que, la función de riesgo se define como

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

y satisface las siguientes condiciones:

$$1. \quad h(t) \geq 0$$

$$2. \quad \int_0^{\infty} h(t) dt = \infty$$

A partir de las definiciones anteriores, es posible obtener relaciones que podrán ser útiles, tales como:

$$S(t) = \exp\left(-\int_0^t h(u) du\right) \quad (3)$$

$$f(t) = h(t) \exp\left(-\int_0^t h(u) du\right)$$

La función de riesgo acumulada, que es una función no negativa y monótona creciente, es definida a través de:

$$H(t) = \int_0^t h(u) du, t \geq 0$$

Donde, por (3), tenemos:

$$S(t) = \exp(-H(t)) \Leftrightarrow H(t) = -\log S(t)$$

esto es, mide el riesgo de ocurrencia del acontecimiento de interés hasta el instante t . (Colosimo, 2008). Tenemos cinco formas posibles que asume la función de riesgo, como se ve en la figura 1. Así esta función puede ser:

1. Monótona creciente: es una forma de la función de riesgo más común en el análisis de sobrevivencia pues corresponde a un riesgo creciente.

2. Monótona decreciente: corresponde a un riesgo decreciente.
3. Constante: ocurre cuando el periodo de observación es corto, o surge una situación imprevista, como enfermedad, accidente y solo es registrado en ese tiempo.
4. Bathtub-shaped: adecuada para caracterizar la mortalidad poblacional, pues en e periodo inicial las muestres resultan esencialmente por enfermedades infantiles, después de que sigue una fase el riego de muestre decrece y se mantiene bajo hasta que ocurra un nuevo aumento debido al envejecimiento.
5. Hump-shaped o unimodal: ocurre, por ejemplo, debido a una intervención quirúrgica, donde el paciente aumenta el riego de muestre en el momento de la operación, pero que se lleva a cabo sin complicaciones

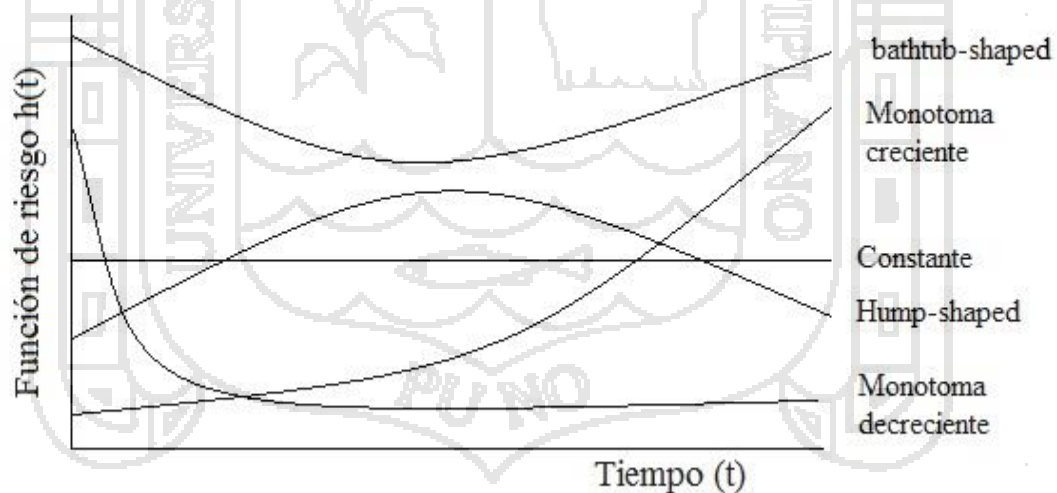


Figura 1. Monotonía de la función de riesgo

2.2 CENSURA

Una característica propia de los datos de sobrevivencia es el hecho de que el

evento de interés no es experimentado en todas las observaciones del estudio. Esta característica es denominada censura. La censura puede admitir dos los límites de tiempo u otra restricción dependiendo de la naturaleza estudio. Se dice que una variable aleatoria es censurada cuando no es posible observar su valor exacto, pero se consigue obtener un límite inferior para ese valor (censura a la derecha), o un límite superior (censura ala izquierda), o ambos (censura intercalar). (Lindsey, 2010)

Existen varios tipos de censuras como ya mencionamos, que pueden ocurrir pero el más común es censura a la derecha.

La censura ocurre a lo debido esencialmente a tres motivos.

- El estudio llego a su fin sin que fuere observado el acontecimiento de interés.
- El estudio individual se perdió durante el seguimiento
- El individuo es retirado del estudio por algún motivo, relacionado con el tiempo de vida.

La figura 2 refleja que puede suceder con los individuos en el estudio, donde Y representa el tiempo de vida de los individuos durante el periodo de observación.

En la primera situación, A, el individuo entra en el inicio del estudio y experimenta el evento, o muerte, en el año 6, lo que significa tener un tiempo de vida de 6 años.

En el caso del individuo B, entra en el inicio del estudio y continúa vivo en el término del mismo, teniendo un tiempo de vida de por lo menos 12 años, esto es, tiene un tiempo censurado de 12 años.

El individuo C, entra más tarde en el estudio, en el 3° año, pero es retirado del estudio en 5° año porque el tratamiento dejo de ser

eficaz, teniendo un tiempo de vida de por lo menos 4 años. El individuo D, no experimenta el evento o muerte, teniendo un tiempo de vida de 6 años. Por último, tenemos el caso del individuo C que es una situación semejante al de C, teniendo un tiempo de vida de por lo menos 8 años, pues a partir de ese año, el individuo no tuvo seguimiento, por algún motivo se perdió el contacto. (Chalita, 2011)

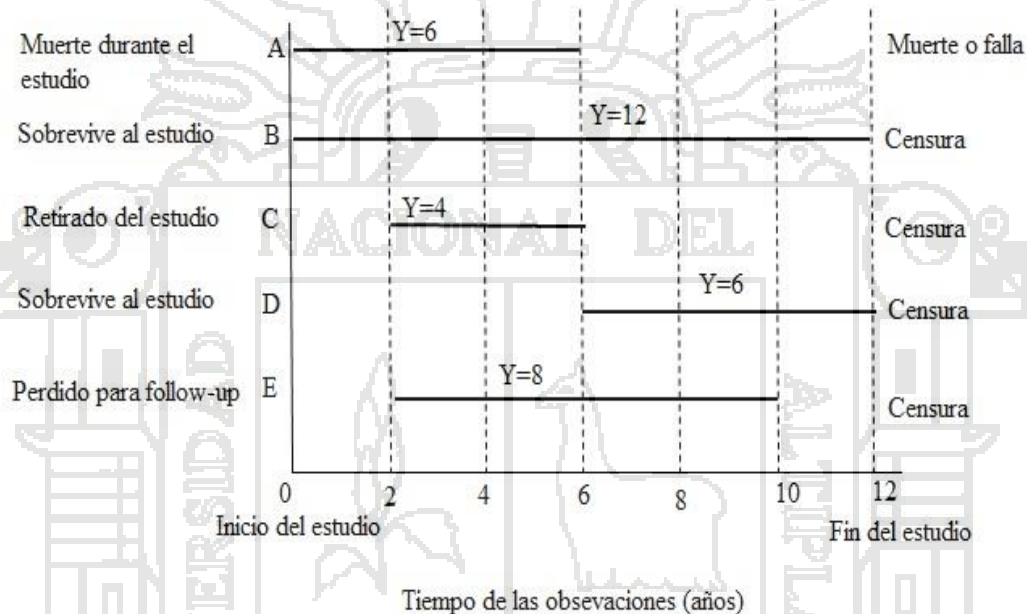


Figura 2. Varios tiempos de censura

Existen esencialmente cinco tipos de censura:

1. **Censura a la derecha:** estamos ante una censura a la derecha cuando apenas sabemos que el tiempo de vida del individuo excede un terminado valor. Puede acontecer por el motivo de que los individuos no quieren permanecer en el estudio y por ello la información queda incompleta, o por que el individuo no experimentó el evento o la muerte antes de finalizar el estudio. Estas situaciones quedaron con información parcial y sabemos que el evento ocurrió

en algún momento después de la fecha de la última visita de seguimiento. A pesar de eso no podemos ignorar esas observaciones ya que no dan alguna información acerca de la sobrevivencia, no sabemos la fecha exacta de su muerte, más sabemos que fue de cierto instante. Podemos dividir este tipo de censura en tres.

- **Tipo I:** este tipo de censura ocurre cuando el estudio es está diseñado para acabar después de x años de seguimiento. En este caso, todos los individuos en que no se hayan observado el acontecimiento de interés durante el estudio, se considera los tiempos censurados en el año x .
 - **Tipo II:** el estudio termina cuando un número de eventos predefinidos ocurre.
 - **Censura aleatoria:** el estudio es diseñando para acabar al final de x años, ósea el individuo entra en el estudio en na fecha de diagnóstico, y debido al hecho de termino del estudio que se fijó previamente, el tiempo que los individuos permanece en el estudio es aleatorio.
2. **Censura izquierda:** Ocurre cuando el tiempo vida es inferior al tiempo observado, es decir, el evento de interés ya ocurrió en algún instante anterior a la observación. Este tipo de censura es menos común.
 3. **Censura intervalar:** Este tipo de censura, sabemos que el evento de interés ocurrió, mas no sabemos cuándo, apenas que ocurrió en un cierto intervalo de tiempo.
 4. **Censura independiente:** Ocurre cuando la razón para ver censura es

independiente de la razón que lleva a la muerte.

5. **Censura no informativa:** En este tipo de censura tenemos que garantizar que el individuo censurado es representativo de todos los individuos que sobreviven y que tienen las mismas características o covariables.

2.3 DISTRIBUCIÓN DE WEIBULL

Para poder establecer un modelo de sobrevivencia es necesario conocer el término de variable aleatoria.

Una variable aleatoria puede concebirse como un valor numérico que está afectado por el azar. Dada una variable aleatoria, no es posible conocer con certeza el valor que esta tomara al ser medida o determinada, aunque si se conoce que existe una distribución de probabilidad asociada al conjunto de valores posibles. (Biswarbrata, 2012)

A continuación, se detallan dos conceptos relativos a la distribución de sobrevivencia de Weibull. (Colosimo, 2008)

La distribución de Weibull fue propuesta por Weibull (1939) y su amplia aplicabilidad fue discutida por este autor. Desde entonces, es usada en estudios biomédicos e industriales. Una variable aleatoria T con distribución Weibull, con parámetro de escala α y parámetro de forma λ , esto es, $T \sim \text{Weibull}(\alpha, \lambda)$ tiene función de densidad:

$$f(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$$

Donde $t > 0$, $\lambda > 0$, $\alpha > 0$. La función de sobrevivencia es:

$$\begin{aligned}
 S(t) &= \int_t^{\infty} \lambda \alpha u^{\alpha-1} \exp(-\lambda u^{\alpha}) du \\
 &= \int_{\lambda t^{\alpha}}^{\infty} e^{-v} dv \left[v = \lambda t^{\alpha} \right] \\
 &= \left[-e^{-v} \right]_{\lambda t^{\alpha}}^{\infty} \\
 &= \exp\left[-\lambda t^{\alpha}\right]
 \end{aligned}$$

entonces

$$S(t) = \exp\{-\lambda t^{\alpha}\}$$

y la función de riesgo es:

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

Una de las grandes características importantes de esta distribución es que presenta una grande variedad de formas, por ejemplo cuando $\alpha = 1$ la función de riesgo es constante y los tiempos de sobrevivencia tienen distribución exponencial con función de densidad $f(t) = \lambda \exp\{-\lambda t\}$.

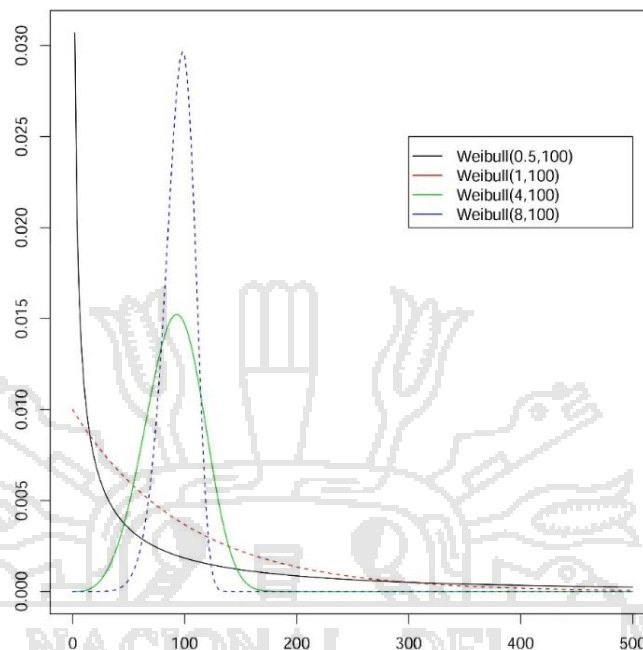


Figura 3. Función de sobrevivencia de la distribución Weibull.

2.4 ANTECEDENTES

(Raony, 2013), llegó a las siguientes conclusiones: Se estimó y comparo curvas de sobrevivencia con censura intervalar informativa, en el cual se presentó un problema cuando utilizamos el tiempo ajustado por la calidad de vida como medida de interés. Se encontró el aumento y disminución de las diferentes curvas de sobrevivencia el cual depende directamente de la calidad de vida y de la forma de cómo es calculado. Se realizó una discusión general sobre el problema de censura informativa en tiempos de sobrevivencia y algunas alternativas en la literatura para hacer inferencia en este contexto. También se enfatizó la importancia de medir la calidad de vida y combinarla con el tiempo de sobrevivencia, metodología que viene ganando espacios en el área de salud la cual carece de métodos y paquetes estadísticos para el análisis.

(Ferreira, 2015), llegó a las siguientes conclusiones: En este estudio no se encontró

evidencia de asociación entre el tratamiento farmacológico y la aparición de exceso de peso, tampoco fue encontrada una diferencia estadística significativa entre la media del peso en los pacientes. Por último es importante hacer referencia que fue posible constatar que efectivamente un elevado porcentaje de los pacientes sometidos a TCPH-lo aumentan de peso sin la influencia aparente de la terapia dada. Sería importante continuar con el estudio prospectivo, en el cual se podría incluir un análisis más

(Carvalho, 2008), llegó a las siguientes conclusiones: Los resultados, tanto bayesiano y clásico, muestran que los valores obtenidos en las estimativas están próximos de los valores reales, lo que sugieren que las estimaciones utilizando ambos métodos de inferencia implementados fueron adecuados. La única reserva se debe cuando la estimación de σ^2_v via Metropolis- Hastings, en el caso de $n_k=15$, existe una sobre estimación y las estimativas no están muy próximas de los valores verdaderos. Con relación a la convergencia de las cadenas, en el aborde bayesiano, los valores de \sqrt{R} están próximos a 1 para todas las estimativas de interés indicando que hubo una convergencia de las cadenas.

(Gallardo, 2014), llegó a las siguientes conclusiones: Se estudió la estimación de los parámetros del modelo vía máxima verosimilitud restringida y ejemplificada a través de estudios de simulación en el que el proceso de estimación fortalece resultados satisfactorios. Además de eso se estudió el impacto en las estimativas considerando erróneamente efectos aleatorios independientes de los efectos aleatorios bivariados y viceversa. Se discutió el hecho de que la inclusión de efectos aleatorios puede ser hecha tanto en el parámetro asociado al número inicial de células cuando el parámetro asociado tiene la probabilidad de activación de cada

célula, pero no en ambos simultáneamente.

(Biswarbrata, 2012), llegó a las siguientes conclusiones: En este trabajo se consideró el método bayesiano y clásico para el análisis de datos con censura intervalar, donde se encontró el estimador de máxima verosimilitud para el modelo de Weibull. El MLE no tiene una forma explícita. El estimador de Bayes bajo un error cuadrado no tiene una forma explícita, entonces se propuso usar técnicas computacionales donde el parámetro de escala es independiente del parámetro a priori.

(Strapasson, 2007), llegó a las siguientes conclusiones: El método M4 (censura intervalar) es la forma correcta para analizar los datos con censura intervalar, considerando los tres tipos de censura con covariable y sin covariable. Cuando la tasa de falla es creciente, el método M2 (punto medio del intervalo) es el mejor método aproximado, considerando los tres tipos de censura con y sin covariable.

(Louzada, 2000), llegó a las siguientes conclusiones: A pesar de la existencia de una mayor flexibilidad en el ajuste de la función de riesgo cuando se considera un procedimiento no paramétrico, en general, esos procedimientos pueden perder precisión. Los estimadores de máxima verosimilitud de los parámetros pueden ser obtenidos numéricamente a través de métodos iterativos, como por ejemplo el método de Newton – Rapson.

(Mendoza Dos Santos, 2013), llegó a las siguientes conclusiones: Se comparó las estimativas del área sobre la curva de ROC considerando estimativas obtenidas a partir de estimadores empíricos de datos sin censura, con la aplicación de la técnica propuesta en este trabajo, censurando parte de las observaciones y excluyendo

las observaciones censuradas. También se observó que un modelo con un número relativamente menor de covariables fue capaz de predecir el tiempo observado, sin el desempeño predictivo del modelo de riesgos proporcionales de Cox estratificado.

(Monteiro, 2014), llegó a las siguientes conclusiones: El programa R presenta ventajas y desventajas, se destaca tres ventajas que son realmente importantes: el hecho de ser gratuito, código abierto, donde se puede crear nuestras propias funciones, se puede modelar funciones que ya existen pero en nuevas situaciones, haciendo del R una herramienta muy versátil y además ocupar poco espacio en la memoria del computador. Las desventajas, son los términos del lenguaje, pero introduciendo paquetes como el *Rcmdr*, el R se torna más amigable.

(Radley, 2012), llegó a las siguientes conclusiones: Intervalos con censura se han convertido cada vez más común en las áreas que producen datos de tiempo de falla. Este tipo de datos viene con frecuencia de las pruebas o situaciones donde los objetos de sus intereses no son constantemente monitoreados. Así los eventos se conocen solamente dentro de un intervalo de tiempo. El propósito de este estudio fue ilustrar métodos paramétricos y no paramétricos disponibles que consideran los datos como intervalo censurado.

CAPÍTULO III

METODOLOGÍA

3.1. ÁMBITO O LUGAR DE ESTUDIO

El lugar de estudio será los pacientes con cáncer de mama de la Clínica Americana-Juliaca 2016.

3.2 .POBLACIÓN Y MUESTRA

La población son todas las personas con cáncer de mama y la muestra será la misma ya que la población es pequeña.

3.3 METODOLOGÍA

Metodología del objetivo: determinar los parámetros de la función de sobrevivencia del modelo de Weibull.

El método de máxima verosimilitud según (Odell, 2010) trata el problema de estimación de la siguiente forma: basado en los resultados obtenidos por la muestra, escogemos la distribución que usaremos, en nuestro caso sería la distribución de Weibull, el método de máxima verosimilitud tiene como finalidad obtener los

estimadores para dichos parámetros, es decir, para α y λ .

Sin embargo, el verdadero valor de t_i no siempre está disponible. En el estudio del cáncer de mama, hay eventos como pacientes que van dejando sus tratamientos, los pacientes mueren por causas diferentes del cáncer de mama, etcétera. Estos acontecimientos causan que el valor de t_i sea censurado, por lo tanto, nos gustaría hacer estimaciones de probabilidad para los datos censurados. Que δ denota el estado de los datos ($\delta = 0$ cuando se censura el tiempo de vida T y $\delta = 1$ cuando se observa el tiempo de vida X). Y T es igual a X si se observa el tiempo de vida, la función de verosimilitud es:

$$l(\theta) \propto \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} [S(t_i, \theta)]^{1-\delta_i}$$

Si asumimos que el dato observado es el siguiente:

$$(T_1, 1), \dots, (T_{n_1}, 1), ([L_{n_1+1}, R_{n_1+1}], 0), \dots, ([L_{n_1+n_2}, R_{n_1+n_2}], 0)$$

Donde n_1, n_2 denotan el número de observaciones no censuras y censuras respectivamente y $n = n_1+n_2$; entonces la función de verosimilitud puede ser escrita como:

$$L(\alpha, \lambda | \text{datos}) = c \alpha^{n_1} \lambda^{n_1} \prod_{i=1}^{n_1} t_i^{\alpha-1} e^{-\lambda \sum_{i=1}^{n_1} t_i^\alpha} \prod_{i=n_1+1}^{n_1+n_2} (e^{-\lambda t_i^\alpha} - e^{-\lambda r_i^\alpha})$$

Donde c es la constante de normalización independiente de λ , la función de log-verosimilitud es:

$$\ln L(\alpha, \lambda | \text{datos}) = l(\alpha, \lambda)$$

$$l(\alpha, \lambda) = \ln c + n_1 \ln \alpha + n_1 \ln \lambda + (\alpha - 1) \sum_{i=1}^{n_1} \ln t_i - \lambda \sum_{i=1}^{n_1} t_i^\alpha + \sum_{i=n_1+1}^{n_1+n_2} \ln(e^{-\lambda t_i^\alpha} - e^{-\lambda r_i^\alpha}) \quad (4)$$

Los estimadores de máxima verosimilitud son obtenidos maximizando la expresión (4), lo que esquivale a resolver el siguiente sistema de ecuaciones no lineales:

$$\frac{\partial l(\alpha, \lambda)}{\partial \alpha} = \frac{n_1}{\alpha} + \sum_{i=1}^{n_1} \ln t_i - \lambda \sum_{i=1}^{n_1} t_i^\alpha \ln t_i + \lambda \sum_{i=n_1+1}^{n_1+n_2} \frac{r_i^\alpha e^{-\lambda r_i^\alpha} \ln r_i - t_i^\alpha e^{-\lambda t_i^\alpha} \ln t_i}{e^{-\lambda t_i^\alpha} - e^{-\lambda r_i^\alpha}} = 0$$

$$\frac{\partial l(\alpha, \lambda)}{\partial \lambda} = \frac{n_1}{\lambda} - \sum_{i=1}^{n_1} t_i^\alpha + \sum_{i=n_1+1}^{n_1+n_2} \frac{r_i^\alpha e^{-\lambda r_i^\alpha} - t_i^\alpha e^{-\lambda t_i^\alpha}}{e^{-\lambda t_i^\alpha} - e^{-\lambda r_i^\alpha}} = 0$$

Como las ecuaciones no son lineales y, en general no poseen soluciones exactas, usamos métodos iterativos, como Newton- Raphson o algoritmo de Quase-Newton. Los métodos iterativos para iniciar el algoritmo necesitan un estimador inicial, el cual será iterado sucesivamente hasta obtener la convergencia

Comparación de funciones de sobrevivencia

Hay numerosas situaciones en las que es necesario comparar dos curvas de sobrevivencia.

Una de las formas más sencillas de llevar a cabo esto, es mediante una visión gráfica.

Sin embargo, existen diversas maneras analíticas de realizar una comparación, como puede ser a través de contrastes de hipótesis basados en tablas de contingencia (Cuadro 1). Donde los grupos 1 y 0 indican cada una de las funciones de sobrevivencia.

Cuadro 1. Tabla de contingencia para el contraste de igualdad de funciones de sobrevivencia en dos grupos en el instante t_i , donde los grupos 1 y 0 indican cada una de las funciones.

Evento	Grupo 1	Grupo 0	Total
Ocurre	$d_1(t_i)$	$d_0(t_i)$	$d(t_i)$
No Ocurre	$n_1(t_i) - d_1(t_i)$	$n_0(t_i) - d_0(t_i)$	$n(t_i) - d(t_i)$
En riesgo	$n_1(t_i)$	$n_0(t_i)$	$n(t_i)$

Para la comparación de estas funciones se construye una prueba de hipótesis, es decir, un procedimiento que permite evaluar hasta qué punto un conjunto de datos observados es consistente con una hipótesis particular, conocida como hipótesis nula H_0 .

En este caso, se toma como H_0 la igualdad de las funciones. De manera que, para construir el estadístico de contraste basta con calcular:

1. El número esperado de ocurrencias de eventos para cada grupo. Por ejemplo, para el grupo 1 es:

$$\hat{e}_1(t_i) = \frac{n_1(t_i)d(t_i)}{n_1(t_i)}$$

2. La varianza estimada del número de ocurrencias de eventos para cada grupo, la cual está basada en la distribución hipergeométrica y para el grupo 1 se define como:

$$\widehat{V}(d_1(t_i)) = \frac{n_1(t_i)n_0(t_i)[n(t_i) - d(t_i)]}{n^2(t_i)[n(t_i) - 1]}$$

Finalmente, el estadístico de contraste para el grupo 1 se define de la siguiente manera.

$$Q = \frac{\left[\sum_{i=1}^m w_i (d_1(t_i) - \widehat{e}_1(t_i)) \right]^2}{\sum_{i=1}^m w_i^2 \widehat{V}(d_1(t_i))}$$

Siendo m el número de tiempos de ocurrencia de eventos en ambos grupos y w_i son los pesos, que varían dependiendo del test utilizado. En el presente trabajo, se utiliza el test log-Rank, en el que se toma $w_i = 1$. Existen otros test diferentes, los cuales pueden consultarse en (Hosmer & Stanley, 1998).

Cuando el número de ocurrencias de eventos es demasiado grande, Q se puede aproximar mediante una distribución Chi-cuadrado χ^2 de un grado de libertad, es decir, $p = P(\chi^2(1) \geq Q)$. Si el valor del p-value p es menor que α (en este estudio, $\alpha = 0,05$), se rechaza H_0 , pues significa que ambas funciones de sobrevivencia son diferentes.

Contraste de hipótesis

Tras el ajuste del modelo de Weibull, se ha de comprobar si las variables del modelo son significativas. Para ello, existen pruebas que se encargan de validar la hipótesis de que los parámetros del modelo de Weibull son asintóticamente equivalentes (Klein, 2000). En ellas, se considera el vector de parámetros estimados

$\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_n)^T$ y la matriz de información evaluada en, definida como:

$$I(\beta) = -\frac{\partial^2 \log(L(\beta))}{\partial \beta_k^2}, \quad k=1, \dots, n$$

A continuación, se describen tres pruebas principales, las cuales son asintóticamente equivalentes, aunque pueden diverger al considerar la muestra estadística.

Test de razón de verosimilitud

Esta prueba es la que presenta una mayor confiabilidad. Sea la hipótesis nula $H_0: \beta = \beta_0$ para la cual la estadística de prueba es

$$X_{LR}^2 = 2[\log(L(\beta_0)) - \log(L(\hat{\beta}))]$$

donde los β_0 son los valores iniciales de los coeficientes y $\hat{\beta}$ es el ajuste obtenido por el modelo de Weibull. Esta prueba sigue la distribución de Chi-cuadrado con n grados de libertad si H_0 es cierta para muestras grandes (Klein, 2000).

Test de Wald

Este test es, quizás, el más natural debido a que se basa en la distribución asintóticamente normal. El estadístico de contraste se define mediante

$$X_w^2 = (\hat{\beta} - \beta_0)^T I^{-1}(\hat{\beta})(\hat{\beta} - \beta_0)$$

donde $I(\beta)$ es la matriz de varianzas y covarianzas estimada.

Esta prueba tiene como hipótesis nula $H_0: \beta = \beta_0$ y sigue la distribución de Chi-cuadrado con n grados de libertad si H_0 es cierta para muestras grandes.

Test de puntajes (score test)

Esta tercera prueba es la conocida como test de los puntajes.

Se define $U(\beta) = (U_1(\beta), U_2(\beta), \dots, U_n(\beta))^T$ como el vector de derivadas de la función de log-verosimilitud parcial, $\log(L(\beta))$. Para muestras grandes, cuando H_0 es cierta, $U(\beta)$ tiene distribución asintótica normal con vector cero por media y matriz de covarianzas dada por $I(\beta)$. Teniendo como hipótesis nula $H_0: \beta = \beta_0$, la estadística de prueba está dada por:

$$X_{sc}^2 = U(\beta_0)^T I^{-1}(\beta_0) U(\beta_0)$$

La cual tiene una distribución Chi-cuadrado con n grados de libertad.

Residuos

Una de las ventajas del enfoque del análisis de sobrevivencia es la posibilidad de efectuar un análisis de residuos (Grambsch, 1990). Un residuo es el valor calculado, para cada paciente, como la diferencia entre el valor de sobrevivencia observado y el valor estimado por la ecuación de regresión. Cuanto mayor es esa diferencia mayor será el valor del residuo, con su signo correspondiente.

El análisis de residuos en cualquier modelo estadístico sirve para verificar la adecuación del modelo ajustado por medio de inspección de gráficos. De manera que, los residuos en el modelo de Weibull pueden ser utilizados para:

- Descubrir la forma funcional correcta de un predictor continuo.
- Identificar los sujetos que están pobremente predichos por el modelo.
- Identificar los puntos o individuos de influencia.
- Verificar el supuesto del modelo de Weibull.

Dentro de este marco estudiaremos los siguientes tipos de residuos.

Residuos de Cox-Snell

Este tipo de residuos, desarrollados por Cox & Snell (Cox & Snell, 1968), sirven para evaluar el ajuste global del modelo planteado. Si el modelo de Weibull definido por es adecuado, entonces las estimaciones del tiempo de sobrevivencia del modelo planteado vienen dadas por un estimador de la función de sobrevivencia $\hat{S}_i(t)$, el cual debe ser muy similar al valor verdadero de $S_i(t)$.

Para evaluar esto, se calculan los residuos de Cox-Snell para los n individuos en estudio del siguiente modo:

$$\hat{e}_i = \hat{\Lambda}(t_i | x_i)$$

Donde $\hat{\Lambda}(\cdot)$ es la función de riesgo acumulado. Para el modelo de Weibull los residuos de Cox-Snell está dado por:

$$\hat{e}_i = (t_i \exp(-x_i' \hat{\beta}))^{\hat{\gamma}}$$

Un resultado importante demostrado por Moeschberger (Klein, 2000), Cox y Snell (Cox & Snell, 1968) y Collet (Collett, 2003), es que, si el modelo apropiado se ajusta bien a los datos, entonces los residuos, tendrán para cada i un valor de distribución exponencial con tasa de riesgo igual a la unidad.

Para probar si los residuos de Cox-Snell están o no aproximadamente distribuidos de forma exponencial, se construye su grafico de residuos (Figura 4). La lógica de este método es sencilla. Si dichos residuos están distribuidos de forma exponencial, entonces una estimación de la tasa de riesgo basada en \hat{e}_i representada frente a

los residuos \hat{e}_i debería tender a una línea recta que pasa por el origen con pendiente igual a la unidad. Es decir, el riesgo acumulado $H(\hat{e}_i)$ frente a los residuos \hat{e}_i , debería ser aproximadamente una línea recta que pasa por el origen con pendiente igual a 1.

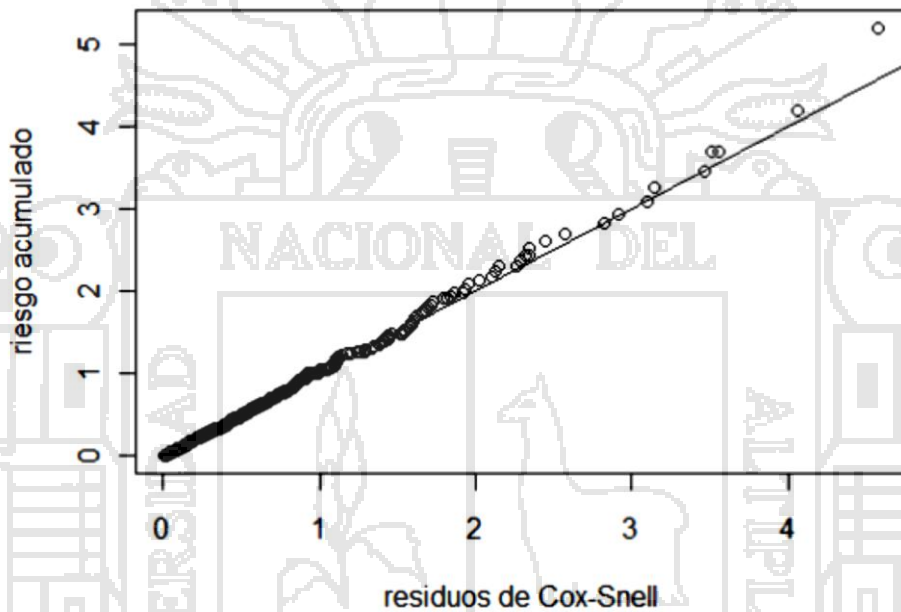


Figura 4. Ejemplo de grafica de residuos de cox-snell, en la cual se representa el riesgo acumulado H_r frente a los residuos r_i .

Residuos de desvíos (deviance)

Los residuos de deviance según (Pruenza, 2014) se obtienen mediante una normalización de los residuos martingala \hat{M}_i y vienen dados por la expresión.

$$d_i = \text{signo}(\hat{M}_i) * \sqrt{-\hat{M}_i - N_i \log\left(\frac{N_i - \hat{M}_i}{N_i}\right)}$$

La transformación de los residuos martingala produce valores simétricos en torno a

cero y, en este caso, el rango de valores de los residuos está comprendido entre $-\infty$ y $+\infty$.

Un residuo con un valor positivo grande corresponderá a pacientes que tienen un tiempo de sobrevivencia grande y los valores estimados por el modelo indican una sobrevivencia menor.

Este tipo de residuos se utiliza para la detección de valores atípicos (outliers). Para ello, se representan los residuos de deviance frente al índice de cada paciente, donde el índice corresponde al número de orden en el que el paciente ha sido registrado en el estudio. Se obtiene una gráfica como la de la Figura 5.

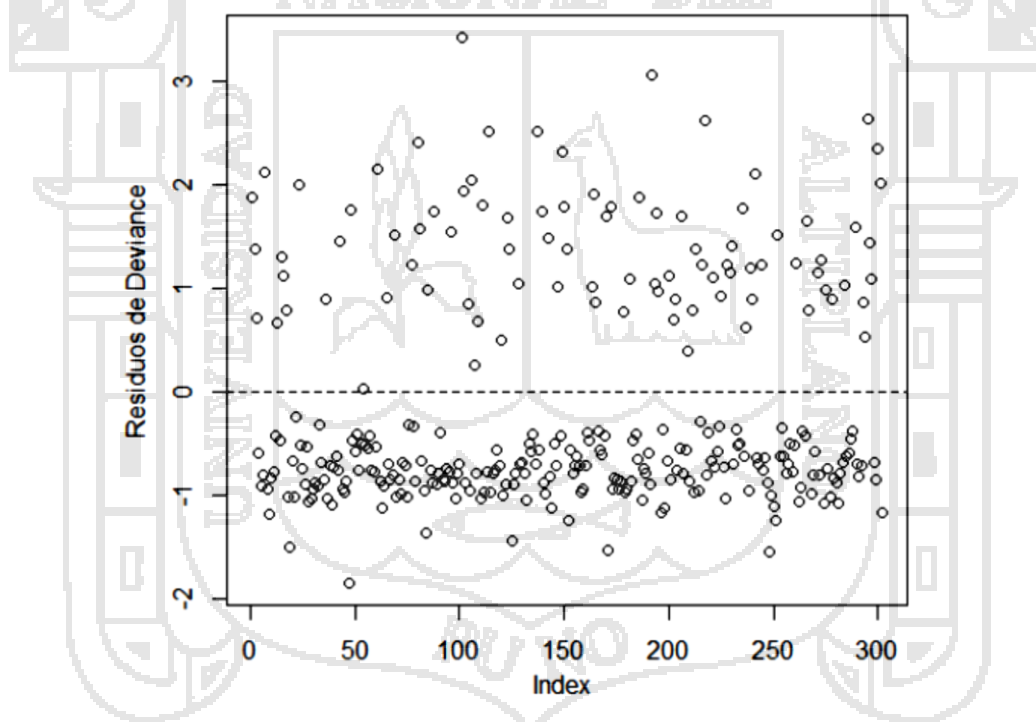


Figura 5. Ejemplo de grafica de residuos de deviance, en la cual se representan los residuos frente al índice de cada paciente.

Aunque en este ejemplo, (Figura 5), no se aprecia ninguna anomalía evidente, en la zona superior aparecen algunas observaciones con residuos positivos grandes

que convendría revisar.

Para detectar dichas anomalías en el ajuste se representa, Figura 5, el valor de los residuos frente al resultado de calcular la puntuación de riesgo para el modelo estimado, es decir, el término $\beta^T X$ donde cada β es el coeficiente estimado y cada X el valor de esa variable para cada paciente.

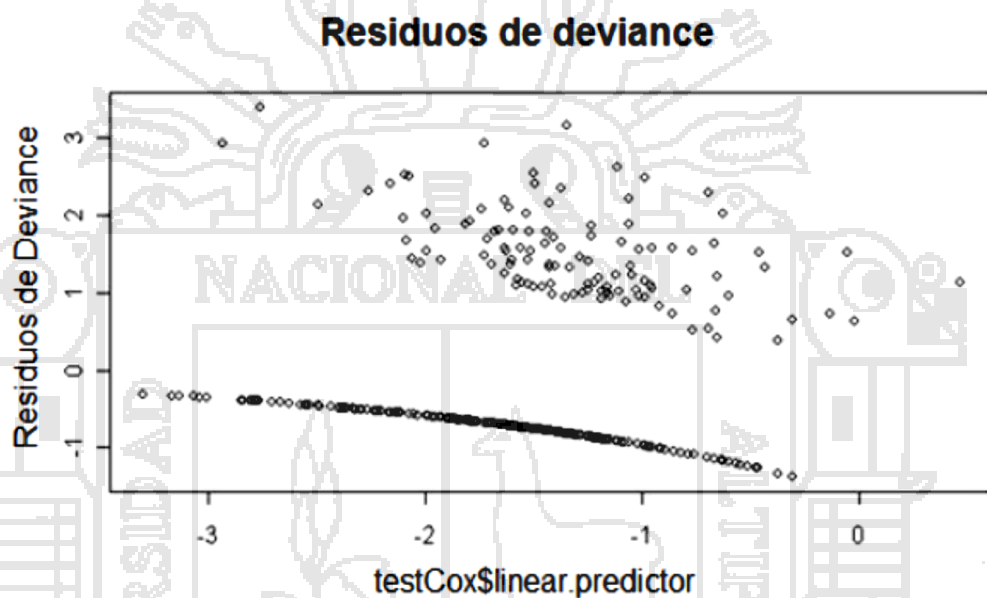


Figura 6. Ejemplo de grafica de residuos de deviance frente a la puntuación de riesgo $\beta^T x$ para cada paciente.

En la Figura 6 se muestra un ejemplo extremo de un modelo en el que se observa un patrón extraño en los residuos que revela que algo no está funcionando en el modelo estimado.

Metodología del objetivo específico: determinar los tiempos con censura intervalar para pacientes con cáncer de mama.

Conjunto de datos

Para la realización del estudio de sobrevivencia, se parte de datos clínicos

previamente normalizados. De esta manera, se consigue que estos sigan una distribución normal y no existan valores demasiados dispersos entre ellos.

El conjunto de datos se divide en dos:

- **Conjunto de entrenamiento (training):** Corresponde al conjunto de datos utilizado para la estimación de los parámetros del modelo. Dicho grupo está formado por un 70% del conjunto inicial.
- **Conjunto de prueba (test):** En este conjunto no intervienen los datos de entrenamiento y, por tanto, se compone por el 30% restante del conjunto de datos de partida. De modo que, una vez construido el modelo, se utiliza este conjunto para evaluarlo y verificar que este se cumple para datos que no jugaron ningún papel en la selección del mismo.

Método de selección de variables

El principal objetivo a la hora de establecer un modelo es buscar aquel que contenga solamente los efectos principales, es decir, con las covariables que han resultado significativas para el modelo de Weibull.

Para determinar cuáles de estas variables aportan más en el modelo y que, a su vez, no estén relacionadas, existen diferentes métodos de selección de variables, los cuales son:

Forward (hacia adelante)

En este método se inicia el proceso con un modelo vacío, solo con el término independiente. Se ajusta un modelo con el método de máxima verosimilitud y se calcula el estadístico Chi-cuadrado con el p-valor de incluir cada variable por

separado. A continuación, se selecciona el modelo con la variable más significativa, es decir, que tiene un p-valor $p < 0,05$.

De nuevo, se ajusta el modelo con las variables seleccionadas y se calcula el p-valor resultante de añadir cada una de las variables que no han sido seleccionadas por separado. Tras escoger el modelo con la variable más significativa, se repiten estos pasos hasta que no queden variables significativas por incluir.

Backward (hacia atrás)

Consiste en empezar con un modelo que contiene todas las variables candidatas e ir eliminando, una a una, cada covariable, a la vez que se calcula la pérdida de ajuste al eliminar.

Se omite del modelo la variable menos significativa, esto es, aquella que mayor p-valor tenga y se repiten estos pasos hasta que todas las variables incluidas sean significativas y no pueda eliminarse ninguna sin que se pierda ajuste.

Stepwise

Este método es una combinación de los procesos forward y backward.

Se puede comenzar, o bien, con el modelo vacío, o bien, con el modelo completo, pero en cada paso se exploran las variables incluidas por si deben salir y las no seleccionadas por si deben entrar en el modelo. Se repiten estos pasos hasta que todas las variables incluidas sean significativas y no entre ni salga ninguna más.

En este trabajo, se utiliza el método backward, por su facilidad a la hora de escoger la variable menos significativa y por permitir ir verificando que el proceso de eliminación es el adecuado.

Tal y como se ha mencionado anteriormente en el documento, se elige como nivel de significación de entrada para los estadísticos un p-valor p menor que $\alpha = 0,05$.

Supongamos que tenemos n individuos bajo estudio y que asociado con el i -ésimo individuo existe un tiempo de vida y un tiempo fijo de censura L_i . Se asume que los T_i son i.i.d con f.d.p. $f(t)$ y función de sobrevivencia $S(t)$. (Lawless, 2003) El tiempo de vida exacto de un individuo puede ser observado como (T_i, δ_i) , donde T_i es el tiempo de falla o de censura, y δ_i es la variable que indica censura o falla, esto es:

$$\delta_i = \begin{cases} 1 & \text{si } T_i \text{ es un tiempo de falla} \\ 0 & \text{si } T_i \text{ es un tiempo censurado} \end{cases}$$

El esquema de censura intervalar puede ser descrito como sigue. Supongamos que n elementos idénticos se ponen en una prueba de vida y conseguimos que T_1, \dots, T_n es la vida útil de estos individuos. Para el elemento i -ésimo, existe un intervalo con censura aleatorizado (L_i, R_i) , que sigue una distribución bivariada. Aquí L_i y R_i denotan el punto final izquierdo y derecha al azar, respectivamente, del intervalo de censura. La vida útil del elemento i -ésimo, T_i , es observable sólo si $T_i \notin [L_i, R_i]$, de lo contrario no es observable. Definimos $\delta_i = I(T_i \notin [L_i, R_i])$, entonces $\delta_i = 1$ implica la observación no es censurada. En este caso se observa el valor real de T_i . Cuando $\delta_i = 0$, sólo se observa el intervalo de censura $[L_i, R_i]$. Para todos los n elementos, los datos observados son de la forma (y_i, δ_i) , $i = 1, \dots, n$, donde

$$(y_i, \delta_i) = \begin{cases} (T_i, 1) & \text{si } T_i \notin [L_i, R_i], \\ ([L_i, R_i], 0) & \text{en otro caso} \end{cases}$$

Para determinar ello usaremos una ficha médica que se encuentra en el anexo 6.



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

En esta sección del documento, se procede a analizar los datos obtenidos tras llevar a cabo un estudio de sobrevivencia.

Para ello, en primer lugar, se realiza una serie de experimentos sobre un toy-example, de manera que sirva para comprobar que los algoritmos implementados y las funciones seleccionadas para llevar a cabo el estudio son las adecuadas y correctas, corrigiendo si fuese necesario algún posible error. A la vez que, se consigue cierto manejo sobre este tipo de datos.

En segundo lugar, se realiza el estudio de sobrevivencia sobre un conjunto de datos reales disponibles de cáncer de mama de la clínica Americana.

Conjunto de datos

El análisis llevado a cabo se basa en un estudio sobre el cáncer de mama con censura intervalar después de haber tenido una cirugía (extracción de la mama), es decir, estaríamos en un nivel de cáncer tipo 2 o 3.

El conjunto de datos consta de 56 personas, información recogida semanalmente, la cual se compone de las siguientes variables explicativas:

Censura: 0- censura, 1- no tiene censura.

Tratamiento: tipo de tratamiento quimioterapia o radioterapia (0- quimioterapia, 1 - radioterapia).

Etapa del cáncer: una versión simplificada de la etapa del cáncer (0- etapa 3,1- etapa 4)

Edad: edad del paciente en el diagnóstico de cáncer de mama.

En este estudio, el conjunto de datos utilizado se compone de 56 pacientes que les ha sido detectado un cáncer de mama. De estos datos se puede destacar que la edad media de los pacientes es 45.44 años.

Cuadro 2. Información sobre el conjunto de datos de cáncer de mama. Se incluye el valor mínimo, máximo, media, desviación típica y mediana de la edad de los pacientes en estudio.

Datos	Edad mínima	Edad máxima	Media	Desv. típica	Mediana
Conjunto completo	31	60	45.44	7.944	46

Para poder realizar un estudio completo sobre la sobrevivencia al cáncer de mama y poder observar la relación que existe entre las distintas variables se elaboran varios experimentos.

4.1 ESTIMACIÓN DE LA FUNCIÓN DE SOBREVIVENCIA

En primer lugar, para realizar el estudio sobre dicho conjunto de datos, se estima la función de sobrevivencia mediante el modelo de Weibull, explicado anteriormente. En la Figura 7, se muestra la función de sobrevivencia obtenida de la estimación sobre el conjunto de datos se puede observar que la probabilidad de sobrevivencia al cáncer de mama disminuye con el tiempo, y en el cuadro 3 presentamos los tiempos con sus respectivas probabilidad, al igual que en el grafico podemos observar que a medida que pasa el tiempo la probabilidad de sobrevivencia disminuye, en (Strapasson, 2007) con el método de Kaplan Meier se determinó que era el método más eficaz para datos censurados, pero aquí vemos el modelo de Weibull es muy preciso para poder encontrar tiempos de sobrevivencia para determinados tiempos

Cuadro 3. Tiempos y probabilidad de sobrevivencia con el modelo de Weibull.

Tiempo (meses)	Probabilidad de sobrevivencia con el Modelo de Weibull
2	0.98841
3	0.9653
4	0.9256
5	0.8675
6	0.7915
7	0.7003
8	0.5988
9	0.4929
10	0.3894
12	0.2119
13	0.1451
14	0.0941
16	0.0333

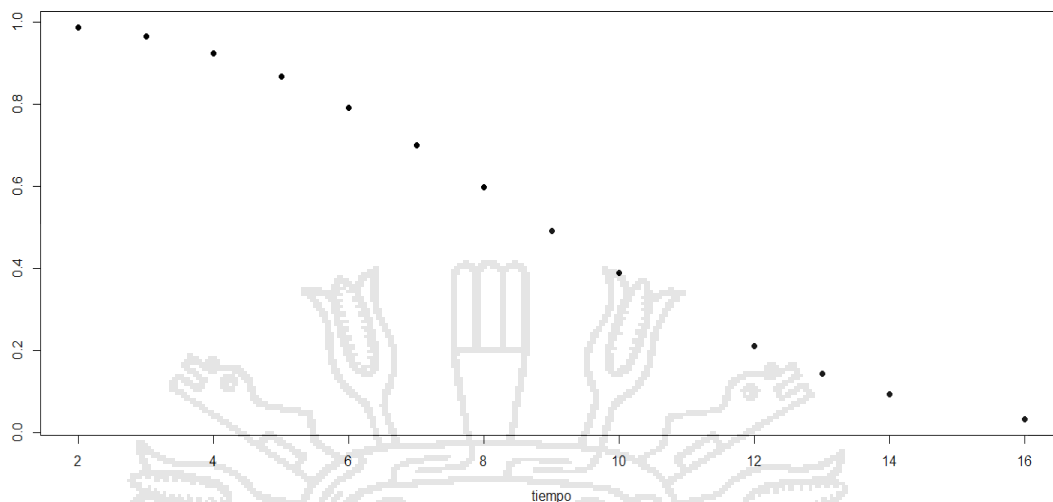


Figura 7. Estimación de Weibull de la función de sobrevivencia.

4.2 COMPARACIÓN DE FUNCIONES DE SOBREVIVENCIA

En esta sección, se comparan las funciones de sobrevivencia de las mujeres que recibieron el tratamiento de quimioterapia y las que recibieron tratamiento de radioterapia.

Para ello construiremos un gráfico, haciendo uso de R, se comprueba que, efectivamente, existe diferencia entre ambas funciones. Obteniendo como resultado la información mostrada en la Cuadro 4 donde se muestran las estimaciones del modelo de Weibull para ambos casos.

Aparentemente, se observa que ambas funciones de sobrevivencia son distintas y que la probabilidad estimada de sobrevivencia de las personas que recibieron tratamiento de quimioterapia tienen una probabilidad de 0.9677 y el grupo que recibió un tratamiento de radioterapia tienen una probabilidad de 0.9048, por lo tanto se puede observar que el tratamiento de quimioterapia tiene mayor probabilidad de sobrevivencia que el tratamiento de radioterapia, siendo el más recomendado para

cáncer de tipo 2 y 3, en (Raony, 2013) se vio que el aumento o disminución de las diferentes curvas de sobrevivencia depende de la calidad de vida y del tratamiento que uno decide tomar, lo cual aquí se verifica que si es necesario utilizar el tratamiento de quimioterapia ya que su probabilidad de vida aumenta.

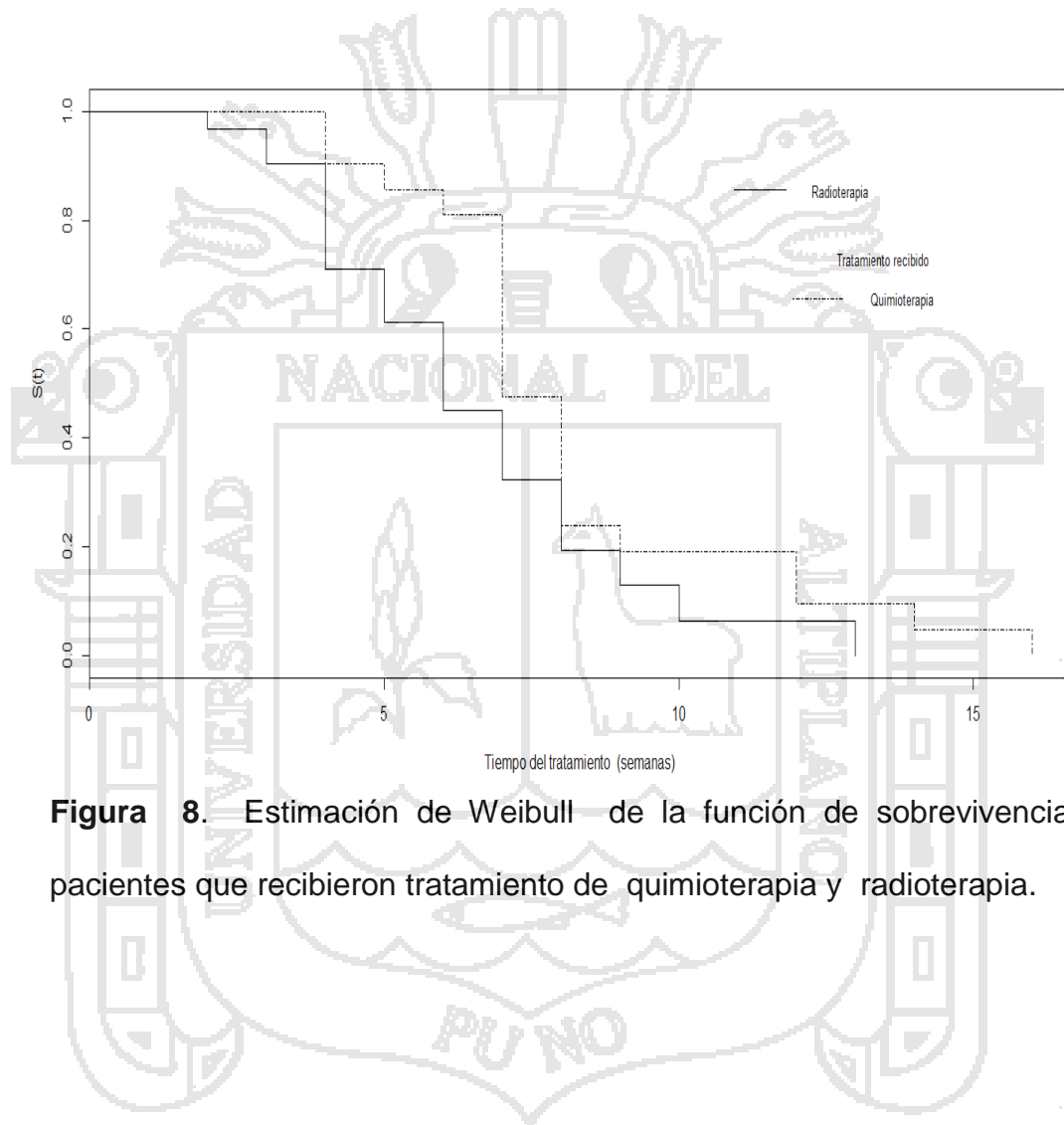


Figura 8. Estimación de Weibull de la función de sobrevivencia para pacientes que recibieron tratamiento de quimioterapia y radioterapia.

Cuadro 4. Resultado de contraste de igualdad de funciones de sobrevivencia para la covariable tratamiento.

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$	Probability
Tratamiento=0	56	21	25.3	1.26	3.39	0.9677
Tratamiento=1	56	35	26.7	1.20	3.39	0.9048

En la cuadro 4, el estadístico de contraste es $Q = 3,39$ lo cual se asocia con un p-value inferior a 0,05. Luego, la hipótesis nula de igualdad de funciones de sobrevivencia (para un nivel de significación del 5 %) se rechaza, lo que indica que hay diferencia entre las funciones de sobrevivencia del grupo de personas que reciben el tratamiento con quimioterapia o radioterapia.

4.3 AJUSTE DEL MODELO DE WEIBULL

Se ajusta una sola vez el modelo de Weibull en el que se incluyen todas las covariables, sin llevar a cabo el método de selección de las mismas. De acuerdo a (Strapasson, 2007), se tiene que mientras utilicemos mas covariables en nuestro modelo, será más óptimo. De manera que, los resultados del ajuste se incluyen en la Cuadro 5.

Cuadro 5. Parámetros del modelo de Weibull para el estudio.

	Coef	se(coef)	Z	Pr(> z)
Trat	0.22557	0.11293	1.997	4.58e-02
Etap	-0.03921	0.11347	-0.346	7.30e-01
Edad	-0.00695	0.00771	-0.902	3.67e-01
Cens	-1.000770	0.10396	-9.693	3.21e-22
Tiempo	0.32409	0.34851	6.669	2.58e-11

La tabla presenta información acerca de las pruebas locales para verificar que cada coeficiente es significativamente distinto de cero. Las columnas proporcionan información para cada covariable como sigue:

Coef: Valor del coeficiente de regresión estimado.

Se (coef): Error estándar del coeficiente de regresión estimado.

z: Corresponde al valor del estadístico, obtenido dividiendo el valor del coeficiente de regresión estimado entre el error estándar estimado.

p: p-value proveniente de una distribución normal con media cero y varianza uno.

Además se obtiene información correspondiente a probar la hipótesis nula de que el vector de variables del modelo son cero, es decir, $H_0: \beta=0$

- **Test de razón de verosimilitud**

La estadística de prueba, denotada anteriormente por X^2_{LR} , es 3,05 y con una

distribución χ^2 con 5 grados de libertad, tiene un p-value $p = 0.003911$, lo que significa que se rechaza la hipótesis nula.

- **Test de Wald**

El test de Wald, denotado anteriormente por X^2_w es 2.89 y con una distribución X^2 con 5 grados de libertad, tiene un p-value $p = 0.004092$. De igual modo, se rechaza la hipótesis nula.

- **Test de puntajes**

La prueba de puntajes, denotada anteriormente por X^2_{sc} , es 2.94 y con una distribución χ^2 con 5 grados de libertad, se obtiene un p-value $p = 0.004014$. Ocurre lo mismo que en los casos anteriores.

Para las tres pruebas anteriores se aprecia un p-value significativamente pequeño (inferior a 0,05) lo cual es evidencia de que, bajo las pruebas realizadas, los coeficientes del modelo son significativamente distintos de cero, y por tanto, se podría considerar que el modelo tiene sentido para las variables explicativas consideradas.

Mediante los valores obtenidos en la Cuadro 5 puede verificar la significación de cada uno de los coeficientes correspondientes a las covariables. De manera que, los coeficientes de *edad*, *trat*, *tiempo*, son significativos al 5 %, y contribuyen significativamente en el modelo planteado y, por tanto, son variables candidatas a ser utilizadas.

Además, los signos de los coeficientes son interpretables, de modo que, un valor positivo para un correspondiente β , significa que el riesgo aumenta con la presencia

de la covariable, y un valor negativo, por el contrario, disminuye el riesgo. Así, la presencia de la covariable E_{tap} ($\beta = -0.03921$) disminuye el riesgo, mientras que ($\beta = 0.32409$) aumenta.

La grafica de la función de sobrevivencia obtenida mediante el modelo de Weibull junto con la obtenida por el estimador de Kaplan-Meier se representa en la Figura 9. Se puede observar que la función de sobrevivencia de Kaplan-Meier es sistemáticamente superior al ajuste del modelo de Weibull, pero ambas prácticamente similares.

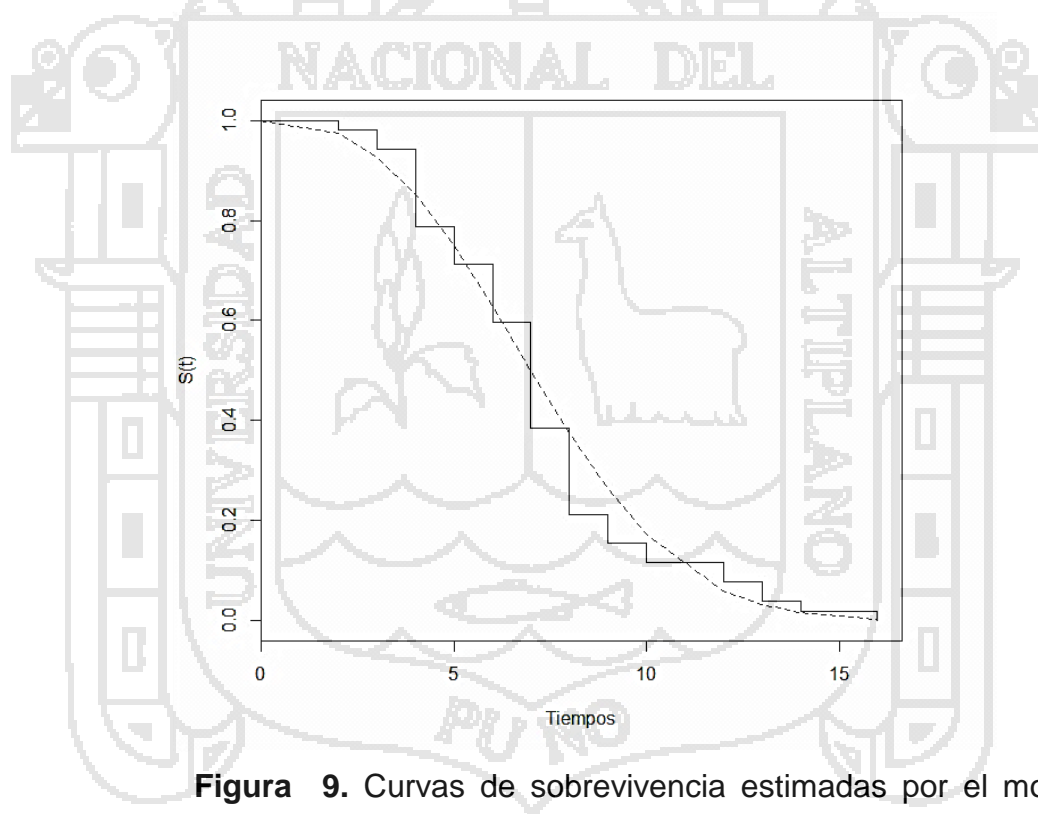


Figura 9. Curvas de sobrevivencia estimadas por el modelo de Weibull versus la curva de sobrevivencia por Kaplan-Meier

4.4 VERIFICACIÓN DEL MODELO DE WEIBULL

Para verificar el modelo de Weibull estimado se realiza un contraste de hipótesis, obteniendo la Cuadro 6. En el que se considera como hipótesis nula el cumplimiento

del supuesto del modelo de Weibull, la cual se rechaza si el valor del p-value es inferior a 0,05.

Cuadro 6 Resultado de contraste del modelo de Weibull para cada covariable se incluye el p-value resultante del contraste.

	P
<i>Trat</i>	0.2323
<i>Etap</i>	0.0738
<i>Edad</i>	0.7992
<i>Cens</i>	0.0657
<i>Tiempo</i>	0.9999

Podemos apreciar que para cada variable el valor de p-value es mayor a 0,05 por lo tanto se considera todas las covariables para nuestro estudio, por lo tanto no existe evidencia significativa al 5% de que se viole el supuesto del modelo de Weibull para las covariables. En (Mendoza Dos Santos, 2013) se encontró que en un modelo con covariables relativamente menor, fue capaz de predecir el tiempo de sobrevivencia, así que en este modelo tenemos 5 covariables, lo cual implica que nuestro modelo será óptimo.

4.5 RESIDUO

4.5.1 Comprobación de la hipótesis global del modelo: residuos de padronizados

Mediante los residuos de Cox-Snell (Cox and Snell, 1968) se trata de evaluar

el ajuste del modelo de Weibull. Si el modelo es correcto y la estimación de los β son cercanas a los valores reales, entonces este estimador deberá describir aproximadamente una recta de pendiente igual a la unidad.

Las gráficas de la Figura 10 indican que, en general, este modelo se ajusta a los datos a pesar de que en algunos casos la estimación de los β se separe de los valores reales. Los valores negativos de los residuos se corresponden con datos censurados. De acuerdo a (Radley , 2012) la propuesta dada sobre modelos paramétricos fue eficaz en este proyecto de tesis, fue muy optimo utilizarlo ya que los datos se ajustan bien. Por lo tanto el modelo de Weibull es óptimo para nuestro estudio de cáncer de mama de personas que residen en la ciudad de Juliaca.

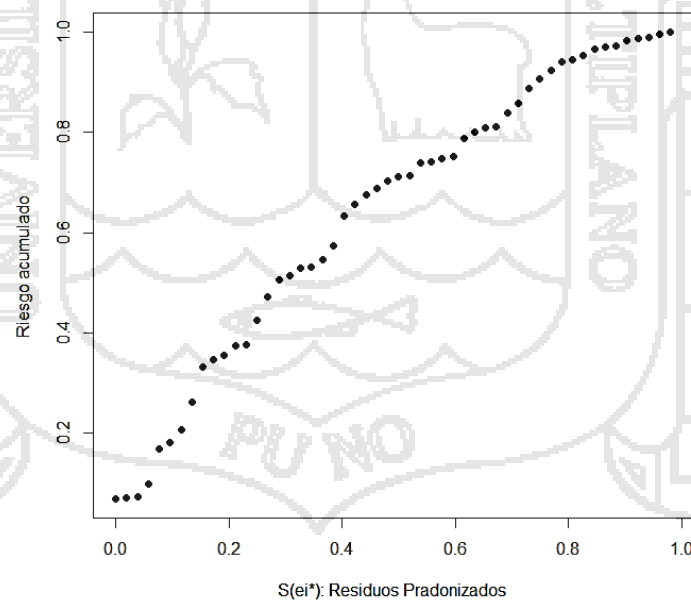


Figura 10 Residuos de cox-snell para datos de cáncer de mama en el estudio de sobrevivencia.

4.5.2 Comprobación de la influencia sobre cada observación en el modelo: residuos de deviance

Los residuos de deviance obtenidos para este estudio se muestran en la Figura 11. En la gráfica, se observa que el número de residuos que está por encima de 1 es superior al que está por debajo de -1, por lo que, hay tendencia a sobrevivir más de lo que predice el modelo. Esto significa que por el método bayesiano propuesto en (Carvalho, 2008) y ahora el con el modelo probabilístico tienen el mismo resultado, siendo este modelo óptimo para estimar la probabilidad de sobrevivencia de personas con cáncer de mama.

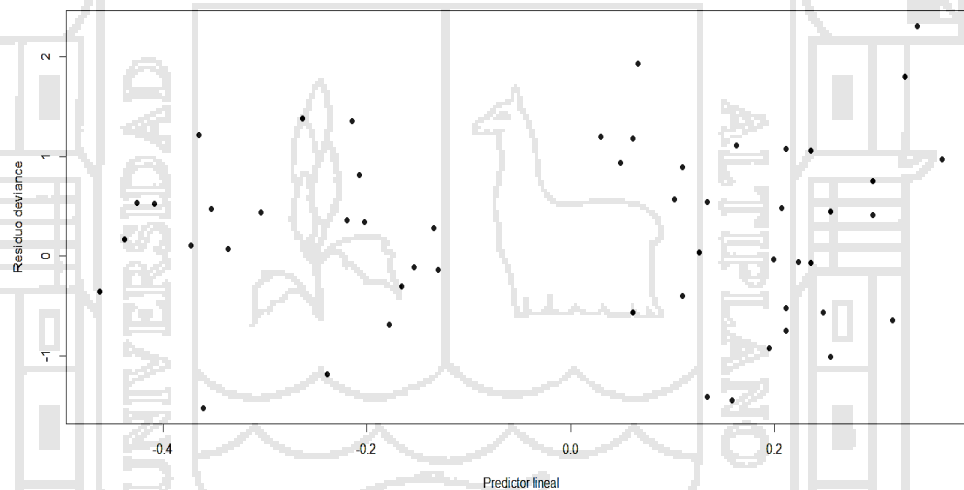


Figura 11. Residuos de deviance para datos de cáncer de mama en el estudio de sobrevivencia para determinar la existencia de outliers en el modelo.

CONCLUSIONES

- Se determinó el modelo probabilístico de Weibull, en cual los tiempos de sobrevivencia está relacionada con el tipo de tratamiento, en donde se demuestra que la quimioterapia tiene una probabilidad de un 96,77% y la radioterapia tiene una probabilidad de un 90,48%, lo cual indica que por cada 100 pacientes, 97 pacientes tienen éxito con el tratamiento de quimioterapia y que por cada 100 pacientes 90 pacientes tienen éxito con el tratamiento de radioterapia, determinándose que la sobrevivencia es más alta con la quimioterapia para pacientes con cáncer de mama de grado 3 y 4.
- Se encontró que los parámetros de la función de sobrevivencia de Weibull de escala es 10.21737 y forma es 2.7392.
- Los tiempos de censura intervalar para pacientes con cáncer de mama para dos meses tiene una probabilidad de 98,84% y para 16 meses una probabilidad de 3,33% de sobrevivencia de pacientes con cáncer de mama grado 3 y 4. Es decir, disminuye la sobrevivencia a medida que aumentan los años que se padece la enfermedad.

- En general, con respecto al estudio de residuos, se puede concluir que el modelo estimado es adecuado y existe una tendencia a sobrevivir más de lo que se predice. Para este fin, se llevó a cabo un estudio con el método de Weibull, analizando las mejores rutinas que el lenguaje R proporciona para ello y desarrollando algunos algoritmos. En general, el modelo de Weibull planteado predice tiempos de sobrevivencia inferiores a los reales, es decir, existe una tendencia a sobrevivir más de lo que el modelo predice.



RECOMENDACIONES

En este se podría realizar un estudio más extenso, incluyendo otros métodos y técnicas que no han sido desarrolladas en este trabajo, como puede ser la curva de ROC.

Como trabajo futuro se recomendaría buscar alternativas en las que se tuviera en cuenta la dependencia temporal de las covariables. Se sugiere el estudio de aproximaciones mediante splines, modelos log-logísticos o el uso de redes neuronales. Cabe destacar que la dificultad en la aplicación de modelos no lineales reside en el tratamiento de datos censurados.

También se podría recomendar un estudio considerando las etapas 1 y 2 del cáncer y cómo influyen los tratamientos de quimioterapia o radioterapia en la probabilidad de supervivencia.

Se debería de considerar dentro de ello más covariables como el tamaño del tumor y el número de ganglios linfáticos afectados ya que esto aumentaría el riesgo de muerte en dichos pacientes.

BIBLIOGRAFÍA

- Biswarbrata, P. (2012). *Analysis of interval- censored data with Weibull Lifetime Distribution*. Sao Paulo: Universidade de Sao Paulo.
- Carvalho, C. M. (2008). *Modelos de sobrevivencia con fracción de cura y efectos aleatorios*. Lisboa: Departamento de estadística e investigación operacional.
- Chalita, L. (2011). *Modelos para datos agrupados e censurados*. New York: Oxford.
- Collett, D. (2003). *Modelling survival data in medical research. 2 da. Edición*. London: Chapman & Hall.
- Colosimo, G. (2008). *Análise de Sobrevivencia aplicada*. Sao Paulo: Edgard Blucher.
- Cox and Snell, D. (1968). *A general definition of residuals*. London: J.R. Statist.
- Ferreira, S. (2015). *Modelos de sobrevivencia para estudios de tiempos hasta la ocurrencia de exceso de peso en individuos adultos*. Lisboa: Departamento de Estadística e Investigación.
- Gallardo, D. (2014). *Extensões em modelos de sobrevivencia com fracao de cura e efeitos aleatorios*. Sao Paulo: Instituto de Matematica e Estadística.

- Grambsch, T. (1990). *Martingale-based residuals for survival models*. London: Biometrika.
- Hosmer & Stanley, D. W. (1998). *Regression Modeling of Time to Event Data*. New York: Wiley Series in Probability and Statistics.
- Klein, J. (2000). *Survival analysis techniques for censored and truncated data*. New York: Springer-Verlang.
- Lawless, J. (2003). *Statistical models and methods for lifetime Data*. 2da Edición. New Jersey: Wiley & Sons.
- Lindsey, R. (2010). *Tutorial in biostatistic methods for interval - censored data*. Chischester.
- Louzada, F. (2000). *Modelos em análise de sobrevivencia*. Sao Paolo: Sao Paolo.
- Mendoza Dos Santos, T. (2013). *Avalizao do desempenho de modelos predictivos no contexto de análise de sobrevivencia*. Sao Paolo: Sao Paolo.
- Monteiro, A. (2014). *Análise de Sobrevivencia en R*. Sao Paolo: Universidade de Sao Paolo.
- Odell, A. P. (2010). *Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model*. Washington.
- Pruenza, C. (2014). *Estudio de análisis de supervivencia*. Madrid: Madrid.
- Radley, S. (2012). *The Statistical Analysis of Interval-Censored Failure Time Data with Applications*. Sao Paolo: Universidade de Paolo.
- Raony, C. (2013). *Estimación y comparación de curvas de sobrevivencia con*

censura informativa. Sao Paolo: Sao Paolo.

Strapasson, E. (2007). *Comparacao de modelos com censura intervalar em análise de sobrevivencia.* Sao Paolo: Sao Paolo.





Anexo 1. Datos referentes al el tiempo de vida para pacientes con cáncer de mama.

de	iz	trat	cens	tiempo	tipo	edad
0	7	1	1	7	0	35
0	8	1	1	8	0	34
0	5	1	1	5	0	40
4	11	1	1	7	0	52
5	12	1	1	7	0	59
5	11	1	1	6	0	53
6	10	1	1	4	1	51
7	16	1	1	9	1	55
7	14	1	1	7	0	45
11	15	1	1	4	1	47
11	18	1	1	7	1	35
17	25	1	1	8	1	58
17	25	1	1	8	1	38
19	35	1	1	16	1	39
18	26	1	1	8	1	41
25	37	1	1	12	0	32
26	40	1	1	14	1	49

27	34	1	1	7	0	41
36	44	1	1	8	1	56
36	48	1	1	12	1	54
37	44	1	1	7	1	52
0	5	0	1	5	0	39
4	9	0	1	5	0	34
4	8	0	1	4	0	35
5	8	0	1	3	1	34
8	12	0	1	4	1	48
8	21	0	1	13	0	41
10	17	0	1	7	1	47
11	13	0	1	2	0	58
11	17	0	1	6	1	53
11	20	0	1	9	1	46
12	20	0	1	8	0	39
14	17	0	1	3	0	57
14	19	0	1	5	1	53
15	22	0	1	7	1	45
16	24	0	1	8	1	46

16	20	0	1	4	1	42
16	24	0	1	8	1	49
17	27	0	1	10	0	51
17	23	0	1	6	0	41
17	26	0	1	9	0	35
18	25	0	1	7	1	39
18	24	0	1	6	1	37
19	32	0	1	13	0	43
22	32	0	1	10	0	46
24	31	0	1	7	1	48
24	30	0	1	6	0	51
30	34	0	1	4	1	46
30	36	0	1	6	0	47
32	40	0	1	8	0	56
35	39	0	1	4	0	60
44	48	0	1	4	1	31

Anexo 2. Programa en R para obtener los parámetros del modelo de Weibull

```
datos1 <-read.table("D:/MAESTRIA/MODELOW/CANCER.txt", header=TRUE)
```

```
attach(datos1)
```

```
ajust2<-survreg(Surv(tiempo,cens)~tipo+edad+trat,dist='weibull')
```

```
ajust2
```

```
alpha<-exp(ajust2$coefficients[1])
```

```
gama<-1/ajust2$scale
```

```
cbind(gama, alpha)
```



Anexo 3. Programa en R para obtener los tiempos de supervivencia y la gráfica en el modelo de Weibull

```

datos1 <- read.table("D:/MAESTRIA/MODELOW/CANCER.txt", header=TRUE)

attach(datos1)

require(survival)

ekm<-survfit(Surv(tiempo,cens)~trat)

st<-ekm$surv

time<-ekm$time

stw<- exp(-(time/10.21737)^2.73)

cbind(time,stw)

summary(ekm)

par(mfrow=c(1,3))

plot(st,stw,pch=16,ylim=range(c(0.0,1)), xlim=range(c(0,1)), xlab = "S(t):
Weibull",ylab="S(t): ")

lines(c(0,1), c(0,1), type="l", lty=1)

par(mfrow=c(1,3))

invst<-qnorm(st)

plot(log(time),log(-log(st)),pch=16,xlab="log(tiempo)",ylab="log(-log(S(t)))")

ajust2$loglik[2]

par(mfrow=c(1,2))

```

```
plot(ekm, conf.int=F, xlab="Tiempos", ylab="S(t)")
```

```
lines(c(0,time),c(1,stw), lty=2)
```

```
legend(25,0.8,lty=c(1,2),c("Kaplan-Meier", "Weibull"),bty="n",cex=0.8)
```



Anexo 4. Programa en R para obtener los tiempos de supervivencia para el tratamiento de quimioterapia y radioterapia.

```
datos1 <- read.table("D:/MAESTRIA/MODELOW/CANCER.txt", header=TRUE)

attach(datos1)

require(survival)

ekm <- survfit(Surv(tiempo,cens)~trat)

summary(ekm)

survdif(Surv(tiempo,cens)~trat,rho=0)

plot(ekm, lty=c(1,4),mark.time=F, xlab="Tiempo del tratamiento
(semanas)",ylab="S(t)")

points(time2, -log(st2))

text(13.5,0.73,c("Tratamiento recibido "),bty="n", cex=0.85)

legend(11.5,0.7,lty=c(4),c("Quimioterapia"),bty="n",cex=0.8)

legend(10.5,0.9,lty=c(1),c("Radioterapia"),bty="n",cex=0.8)
```


Anexo 5. Programa en R para obtener los residuos padronizados y de deviance

```

datos1 <-read.table("D:/MAESTRIA/MODELOW/CANCER.txt", header=TRUE)

attach(datos1)

require(survival)

fit3<-coxph(Surv(tiempo,cens)~tipo+edad+trat,data=datos1,x
T,method="weibull")

summary(fit3)

-2*fit3$loglik[2]

resid(fit3,type="scaledsch")

cox.zph(fit3, transform="identity") ## g(t) = t

par(mfrow=c(2,3))

plot(cox.zph(fit3))

par(mfrow=c(1,2))

rd<-resid(fit3,type="deviance") # residuos deviance

rm<-resid(fit3,type="martingale") # residuos martingal

pl<-fit3$linear.predictors

plot(pl,rm, xlab="Preditor linear", ylab="Resíduo martingal", pch=16)

plot(pl,rd, xlab="Preditor linear", ylab="Resíduo deviance", pch=16)

xb<-ajust1$coefficients[1]+ajust1$coefficients[2]*tipo+ajust1$coefficients[3]*edad+

ajust1$coefficients[4]*trat

```

```
sigma<-ajust1$scale

res<-((log(tiempo)-(xb))/sigma)          # residuos padronizados

resid<-exp(res)                          # exponencial dos residuos padronizados

ekm<- survfit(Surv(resid,cens)~1)

resid<-ekm$time

sln<-pnorm(-log(resid))

par(mfrow=c(1,2))

plot(ekm$surv,sln, xlab="S(ei*): Residuos Pradonizados",ylab="Riesgo acumulado",pch=16)

plot(ekm,conf.int=F,mark.time=F,xlab="Residuos",ylab="Sobrevivencia",pch=16)

lines(resid,sln,lty=2)

legend(1.3,0.8,lty=c(1,2),c("Residuos","Weibull"),cex=0.8,bty="n")
```

Anexo 6. Ficha de visita médica

APELLIDOS Y NOMBRES			C.I. N°		
DIRECCION DE DOMICILIO			TELEFONO		
LUGAR DE NACIMIENTO	EDAD	SEXO	PESO	TALLA	SUP. CORPORAL

TRATAMIENTO

DATOS GENERALES

--

ESTATUS CONSTITUCIONAL

GRADO 0:			GRADO 1:	GRADO 2:	GRADO 3:	GRADO 4:
FECHA DE INICIO TRATAMIENTO		HORA	N° DE CICLOS APLICADOS	N° CICLOS PENDIENTES		
DIA	MES	AÑO				

APLICACIÓN

INTERVALO	HOSPITAL CLINICA	O	ENTIDAD FEDERAL