



# Introducción al Análisis de Supervivencia

Sergio M. Nava Muñoz  
nava@cimat.mx  
CIMAT  
2025-06-01

# Análisis de Supervivencia

## Parte 1: Introducción y objetivos

### Introducción y Objetivos

- **Relevancia de la Supervivencia** en contextos biomédicos, industriales y económicos
- **Objetivos de la sesión:**
  - Comprender conceptos clave: tiempo de falla, funciones de supervivencia y riesgo
  - Mecanismos de censura y truncamiento
  - Modelos de supervivencia básicos y estimación en R
  - Aplicaciones prácticas con ejemplos de datos reales

### Agenda

1. Definición y ejemplos de tiempos de falla
2. Función de supervivencia y función de riesgo
3. Censura y truncamiento
4. Estimación no paramétrica: Kaplan-Meier
5. Modelos paramétricos

### 1. ¿Qué es el análisis de supervivencia?

**Análisis de Supervivencia** es el análisis estadístico de tiempos hasta la **ocurrencia de un evento**. También se conoce como análisis de *time-to-event*.

Más precisamente, estudia el **tiempo transcurrido entre dos eventos**:

- Un evento de **inicio**
- Un evento de **fin**

**Símbolo usual:**  $T$   
→ Variable aleatoria no negativa que representa el tiempo hasta el evento.

**Términos comunes:**

- Tiempo de falla
- Tiempo de vida
- Tiempo de supervivencia



## ¿Dónde se aplica el análisis de supervivencia?

### Aplicaciones comunes

- **Biomédicas**
  - Tiempo hasta recaída, muerte, recuperación o aparición de una enfermedad.
- **Industriales**
  - Duración de dispositivos, tiempo hasta la primera falla.
- **Económicas / financieras**
  - Tiempo en desempleo, tiempo hasta bancarrota, duración de relaciones laborales.

**Nota:** El "tiempo" puede medirse en días, semanas, kilómetros, horas de operación, etc.



## ¿Qué se necesita definir para analizar tiempos de falla?

Para interpretar adecuadamente los datos, es necesario:

1. **Evento de origen**
  - ¿Desde cuándo empieza a contarse el tiempo?
2. **Escala de medición**
  - ¿Cómo se mide el tiempo? (reloj, kilómetros, ciclos)
3. **Evento de fin**
  - ¿Qué se considera "fallo", "recuperación", o "evento"?



## Ejemplos de datos de supervivencia

En la continuación se presentan algunos ejemplos de datos de supervivencia. Estos ejemplos fueron obtenidos de Klein & Moeschberger (1997).

### Ejemplo de Duración de la remisión en un ensayo clínico para leucemia aguda

Duración de remisión de un ensayo clínico para leucemia aguda. Resultados de un ensayo clínico en donde se quería comparar la efectividad de la droga 6-MP versus placebo en 42 niños con leucemia aguda. El evento de inicio es remisión parcial de la enfermedad después de haber sido tratados con la droga prednisona. El evento de fin es recaída o muerte. La escala de medición es tiempo calendario en meses. Algunos individuos no presentaron el evento de fin al término del estudio. Estos casos son marcados con un + y son llamados censurados por la derecha. Más adelante los veremos con detalle.

TABLE 1.1 Remission duration of 6-MP versus placebo in children with acute leukemia			
Pair	Remission Status at Randomization	Time to Relapse for Placebo Patients	Time to Relapse for 6-MP Patients
1	Partial Remission	1	10
2	Complete Remission	22	7
3	Complete Remission	3	32*
4	Complete Remission	12	25
5	Complete Remission	8	22
6	Partial Remission	17	6
7	Complete Remission	2	16
8	Complete Remission	11	34*
9	Complete Remission	8	32*
10	Complete Remission	12	25*
11	Complete Remission	2	11*
12	Partial Remission	5	20*
13	Complete Remission	4	19*
14	Complete Remission	15	6
15	Complete Remission	8	17*
16	Partial Remission	23	35*
17	Partial Remission	5	6
18	Complete Remission	11	15
19	Complete Remission	4	9*
20	Complete Remission	1	6*
21	Complete Remission	8	10*

\* Censored observation

7



## Naturaleza de los datos de supervivencia

- Los datos de tiempo a evento son realizaciones de **variables aleatorias no negativas**.
- Pueden ser:
  - **Continuos:** como tiempo en días, semanas, horas.
  - **Discretos:** como número de visitas o ciclos.

### Ejemplo de variable de interés:

$T$  = Tiempo entre el ingreso al hospital y la recuperación

9



## Ejemplos prácticos

### Ejemplo biomédico – Ensayo clínico:

- Evento de origen: entrada del paciente al estudio
- Evento de fin: muerte o recuperación
- Escala: tiempo en semanas

### Ejemplo industrial – Billetes:

- Evento de origen: salida a circulación
- Evento de fin: destrucción por deterioro
- Escala: tiempo calendario o número de transacciones

### Ejemplo financiero – Desempleo:

- Evento de origen: pérdida de empleo
- Evento de fin: contratación nueva
- Escala: meses sin empleo

11



## Ejemplo de Transplante de médula ósea en pacientes con leucemia.

El transplante de médula es un procedimiento estándar en pacientes con leucemia aguda. La recuperación después del transplante es un proceso complejo. La prognosis para la recuperación puede depender de factores que se conocen al momento del transplante, como edad y sexo del paciente y donador, etapa de la enfermedad inicial, tiempo entre el diagnóstico y el transplante, etc. La prognosis final depende de cómo evoluciona el paciente después del transplante. Puede generar aversión o rechazo de la médula transplantada (GVHD), que el conteo de plaquetas se vuelva normal o desarrollar infecciones, etc. El transplante se considera fracaso cuando el paciente recae o muere.

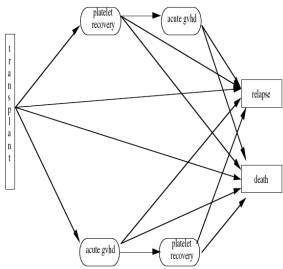


Figure 1.3 Recovery Process from a Bone Marrow Transplant

8

10

12

## Ejemplo de Transplante de médula ósea en pacientes con leucemia. (cont.)

CIMAT

Table 0.1

(continued)

$g$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	$x_{18}$	$x_{19}$	$x_{20}$	
1	2081	2081	0	0	0	67	1	121	1	13	1	26	33	1	0	1	98	0	1	0	
1	1602	1602	0	0	0	9602	0	139	1	10	1	21	37	1	1	0	1720	0	1	0	
1	1496	1496	0	0	0	1096	0	307	1	12	1	26	35	1	1	1	0	127	0	1	0
1	1452	1452	0	0	0	70	1	95	1	12	1	17	21	0	1	0	108	0	1	0	
1	1435	1435	0	0	0	1435	0	250	1	12	1	52	50	1	1	1	93	0	1	0	
1	1177	1177	0	0	0	1375	0	135	1	12	1	21	35	1	1	1	1007	0	1	0	
1	1330	1330	0	0	0	1330	0	96	1	17	1	20	17	1	0	1	1005	0	1	0	
1	906	906	0	0	0	72	1	131	1	12	1	21	24	1	0	5	1331	0	1	0	
1	226	226	0	0	0	226	0	240	0	10	1	18	23	0	1	0	208	0	1	0	
1	1109	1109	0	0	0	1109	0	70	1	20	1	24	40	1	1	0	174	0	3	1	
1	1111	1111	0	0	0	1111	0	1111	0	22	1	19	28	1	1	0	250	0	3	1	
1	530	530	0	0	0	38	1	94	1	34	1	17	20	1	1	0	131	0	3	1	

TABLE 0.1  
Data on 137 Bone Marrow Transplant Patients

$g$ —Disease group  
1—ALL  
2—AML low-risk  
3—AML high-risk  
 $T_0$ —Time (in days) to death or on study time  
 $T_1$ —Disease-Free survival time (time to relapse, death or end of study)  
 $R_0$ —Death indicator  
1—Dead  
0—Alive  
 $R_1$ —Relapse indicator  
1—Relapsed  
0—Disease-Free  
 $R_2$ —Disease-Free survival indicator  
1—Dead or relapsed  
0—Alive disease-free  
 $T_2$ —Time (in days) to acute graft-versus-host disease  
 $R_3$ —Acute graft-versus-host disease indicator  
1—Developed acute graft-versus-host disease  
0—Never developed acute graft-versus-host disease  
 $T_3$ —Time (in days) to chronic graft-versus-host disease  
 $R_4$ —Chronic graft-versus-host disease indicator  
1—Developed chronic graft-versus-host disease  
0—Never developed chronic graft-versus-host disease  
 $T_4$ —Time (in days) to return of platelets to normal levels  
 $R_5$ —Platelet recovery indicator  
1—Platelets returned to normal levels  
0—Platelets never returned to normal levels  
 $Z_1$ —Patient age in years  
 $Z_2$ —Donor age in years  
 $Z_3$ —Patient sex  
1—Male  
0—Female  
 $Z_4$ —Donor sex  
1—Male  
0—Female  
 $Z_5$ —Patient CMV status  
1—CMV positive  
0—CMV negative  
 $Z_6$ —Donor CMV status  
1—CMV positive  
0—CMV negative  
 $Z_7$ —Waiting time to transplant in days  
 $Z_8$ —Hb  
1—Hb Grade 4 Or 5 and AML  
0—Otherwise  
 $Z_9$ —Hsped  
1—The Ohio State University  
2—Alfred  
3—St. Vincent  
4—Hahnemann  
 $Z_{10}$ —MFC used as a graft-versus-host prophylactic  
1—Yes  
0—No

## Ejemplo Tiempo al primer uso de marihuana.

CIMAT En este estudio a 191 estudiantes de preparatoria se les preguntó: ¿Cuál fue la primera vez que probaste la marihuana?. Las respuestas fueron, "la edad exacta a la que la probaron", "nunca la he probado", y "la probé pero no recuerdo cuando fue la primera vez". En este último caso tenemos una censura por la izquierda. El evento de interés ha ocurrido en algún momento previo a la edad actual del estudiante!.

TABLE 1.8  
Marijuana use in high school boys

Age	Number of Exact Observations	Number Who Have Yet to Smoke Marijuana	Number Who Have Started Smoking at an Earlier Age
10	4	3	0
11	12	3	0
12	19	2	0
13	24	15	1
14	20	24	2
15	13	18	3
16	5	14	2
17	1	5	3
18	0	3	1
>18	4	3	0

Infection Time	Adult Induction Time	Child Induction Time
4.00	1, 1.5, ..., 4	1
4.25	1.25, 1.5, ..., 3.5	1.75
4.50	1, 1.5, ..., 3.25	3.25
4.75	1, 1.5, ..., 3.25	1, 2.25
5.00	0.5, 1.5, ..., 3	0.5, 0.75, 1.5, 2.5
5.25	0.25, 0.25, ..., 2.75	0.25, 1, 1.5
5.50	1, 1, ..., 2.5	0.5, 1.5, 2.5
5.75	0.25, 0.75, ..., 2.25	1.75
6.00	0.5, 0.75, ..., 2	0.5, 1.25
6.25	0.75, 1, ..., 1.75	0.5, 1.25
6.50	0.25, 0.25, ..., 1.5	0.75
6.75	0.75, 0.75, ..., 1.25	0.5, 0.75
7.00	0.75	0.75
7.25	0.25	0.25

CIMAT

## Ejemplo Tiempos de muerte de adultos mayores residentes de un asilo.

CIMAT

Channing House es una casa de retiro en California. Datos con las edades de muerte de 462 individuos (97 hombres y 365 mujeres) que estuvieron en la residencia durante el periodo de enero de 1964 y julio de 1975. Se reportó la edad a la muerte o al momento en que se salían del asilo (en meses) y la edad a la que los individuos entraron al asilo. Estos datos son un ejemplo de truncamiento por la izquierda que más adelante veremos con detalle. Un individuo tiene que sobrevivir lo suficiente para estar en edad de entrar al asilo. Individuos que mueren previamente a la edad de retiro son excluidos del estudio.

Observation number  
Death status (1=dead, 0=alive)  
Age of entry into retirement home, months  
Age of death or left retirement home, months  
Difference between the above two ages, months  
Gender (1=male, 2=female)  
Reference: Hyde Biometrika (1977), 225-230.  
THE DATA  
1 1 1042 1172 130 2  
2 1 921 1040 119 2  
3 1 895 1003 118 2  
4 1 901 1018 117 2  
5 1 808 932 124 2  
6 1 915 1004 89 2  
7 1 901 1023 122 2  
8 1 852 908 56 2  
9 1 828 868 40 2  
10 1 968 990 22 2

15

## Tiempo a desarrollar sida.

CIMAT

Se reportan datos con tiempos de infección y de inducción para 258 adultos y 37 niños que fueron infectados con el virus del VIH y desarrollaron sida antes del 30 de junio de 1986. Los datos consisten de los tiempos (en años) desde que adultos fueron infectados por el virus por transfusión de sangre contaminada, y el tiempo de espera hasta el desarrollo de sida. Para la población pediátrica, los niños fueron infectados in útero o al nacer. El tiempo base de medición es el 1 de abril de 1978. En este estudio, sólo los individuos que han desarrollado sida antes del término del estudio son considerados. Individuos que no han desarrollado sida no son incluidos en el estudio. Este tipo de datos es llamado truncados por la derecha y más adelante los veremos con detalle.

### Tiempo de inducción de SIDA en adultos y niños

Infection Time	Adult Induction Time	Child Induction Time
0.00	5	
0.25	6.75	
0.75	5, 5, 7.25	
1.00	4.25, 5.75, 6.25, 6.5	5.5
1.25	4, 4.25, 4.75, 5.75	
1.50	2.75, 3.75, 5, 5.5, 6.5	2.25
1.75	2.75, 3, 5.25, 5.25	
2.00	2.25, 3, 4, 4.5, 4.75, 5, ...	
2.25	3, 5.5	3
2.50	2.25, 2.25, ..., 4	
2.75	1.25, 1.5, ..., 5.25	1
3.00	2, 3.25, ..., 5	1.75
3.25	1.25, 1.75, ..., 4.5	
3.50	1.25, 2.25, ..., 4.5	0.75
3.75	1.25, 1.75, ..., 4.25	0.75, 1, ..., 4.25

## Ejemplos de Datos de Supervivencia

CIMAT

```
# Ejemplo simulado de tiempos de supervivencia
2 set.seed(123)
3 tiempos <- rexp(8, rate = 0.05)
4 status <- rbinom(8, 1, prob = 0.8)
5 library(survival)
6 data_sim <- data.frame(time = tiempos, event = status)
```

Primeros 8 registros  
de data\_sim

time	event
16.8691452	1
11.5322054	0
26.5810974	1
0.6315472	1
1.1242195	1
6.3300243	0
6.2845458	0
2.9053361	1

14

16

17

## Parte 2: Conceptos de Análisis de Supervivencia

### Funciones fundamentales

Antes de introducir las funciones de supervivencia y riesgo, recordemos dos funciones clave en probabilidad:

- **Función de densidad:**  $f(t)$
- **Función de distribución acumulada:**  $F(t) = P(T \leq t)$

### Función de distribución acumulada $F(t)$

- Es la probabilidad de que la variable aleatoria tome un valor menor o igual que  $t$ :

$$F(t) = \int_{-\infty}^t f(u) du = P(T \leq t)$$

- **Propiedades:**
  - $F(t)$  es monótona creciente
  - $\lim_{t \rightarrow -\infty} F(t) = 0$
  - $\lim_{t \rightarrow \infty} F(t) = 1$

18

### Introducción

En esta sección abordaremos los conceptos fundamentales para el análisis de datos de supervivencia, comenzando con funciones de probabilidad clásicas y avanzando hacia funciones específicas como la función de supervivencia y la función de riesgo.

#### Objetivos

- Recordar las funciones de densidad y distribución acumulada.
- Introducir la función de supervivencia  $S(t)$  y la función de riesgo  $h(t)$ .
- Interpretar estas funciones desde una perspectiva probabilística.
- Visualizar ejemplos aplicados y comparativos con distintas distribuciones.

20

### Función de densidad $f(t)$

- Describe la distribución de probabilidad de una variable continua  $T$
- No es una probabilidad en sí, pero su integral sí lo es:

$$P(a < T \leq b) = \int_a^b f(t) dt$$

- Debe cumplir:

$$f(t) \geq 0 \quad \text{y} \quad \int_{-\infty}^{\infty} f(t) dt = 1$$

22

### Relación entre $f(t)$ y $F(t)$

- Si  $f$  es continua:

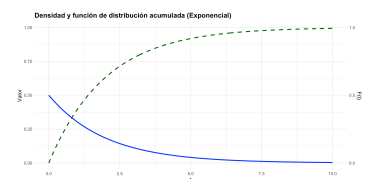
$$f(t) = \frac{d}{dt} F(t)$$

- Y también:

$$F(t) = \int_{-\infty}^t f(u) du$$

Estas relaciones son clave para definir funciones como la de supervivencia y la de riesgo, que veremos a continuación.

### Ejemplo en R: distribución exponencial con parámetro $\lambda = 0.5$



21

23

## Funciones fundamentales en análisis de supervivencia

En análisis de supervivencia, las variables aleatorias de interés  $T$  son no negativas, y se caracterizan no solo por  $f(t)$  o  $F(t)$ , sino también por funciones **más interpretables**:

- $S(t)$ : función de supervivencia
- $h(t)$ : función de riesgo o tasa de falla
- $H(t)$ : riesgo acumulado

## Parte 3: Función de supervivencia $S(t)$

### Función de Supervivencia

$$S(t) = P(T > t) = 1 - F(t)$$

Representa la probabilidad de **sobrevivir más allá del tiempo  $t$** .

**Propiedades clave:**

- Monótona no creciente
- $S(0) = 1, \lim_{t \rightarrow \infty} S(t) = 0$

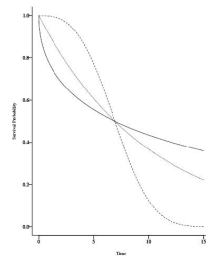
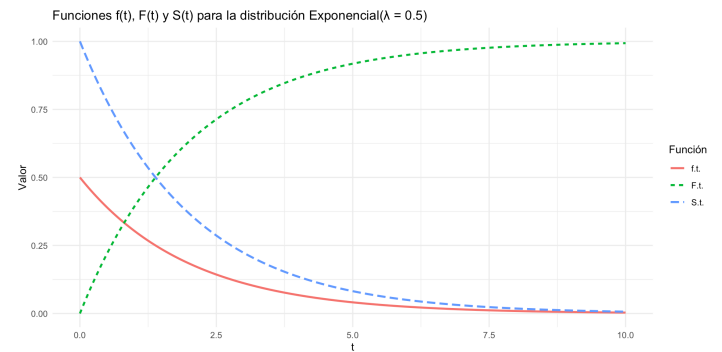


Figure 2.1 Weibull Survival function for  $\alpha = 0.5, \lambda = 0.26528$  (—);  $\alpha = 1.0, \lambda = 0.1$  (---);  $\alpha = 3.0, \lambda = 0.00208$  (.....).

### Ejemplo: función de supervivencia para distribución exponencial

Sea  $T \sim \text{Exp}(\lambda = 0.5)$ , es decir:

$$f(t) = \lambda e^{-\lambda t}, \quad F(t) = 1 - e^{-\lambda t}, \quad S(t) = e^{-\lambda t}$$



## Parte 4: Función de riesgo $h(t)$

### Función de Riesgo

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

- También conocida como:
  - Tasa de falla condicional (confiabilidad)
  - Tasa de mortalidad (demografía)
  - Función de intensidad (procesos estocásticos)

**Interpretación:**

Tasa instantánea de ocurrencia del evento, dado que se ha sobrevivido hasta  $t$ .

## Ejemplos de formas de riesgo

Forma del riesgo	Interpretación
Riesgo creciente	Envejecimiento
Riesgo decreciente	Rejuvenecimiento
Riesgo tipo "tina de baño"	Mortalidad neonatal y senil
Riesgo tipo "montaña"	Recada tras tratamiento

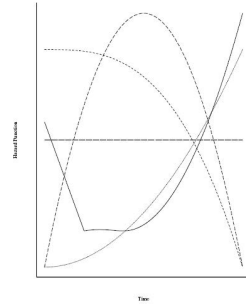
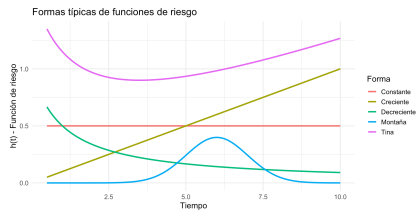


Figure 2.4 Shapes of hazard functions. Constant hazard (—), increasing hazard (—), decreasing hazard (---), bathtub shaped (—), unimodal (—).

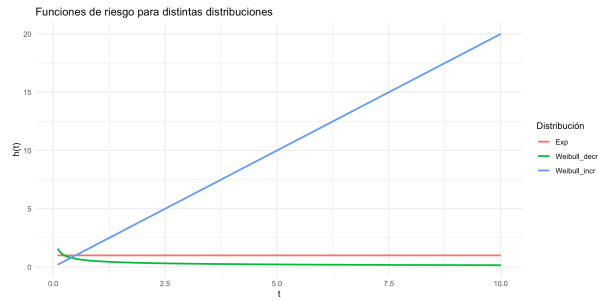
#

## Ejemplo: función de riesgo para distribuciones comunes

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

Para la distribución exponencial con  $\lambda = 0.5$ ,  $h(t) = \lambda$ , constante.

Comparémosla con la distribución Weibull, donde el riesgo puede aumentar o disminuir con el tiempo.



## Otra forma de visualización

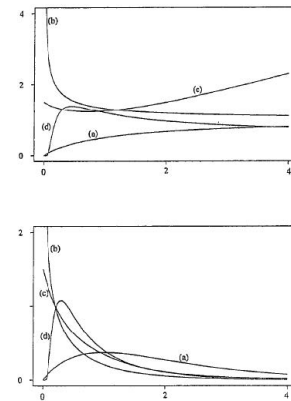


Figure 1.1. Some hazard and probability density functions.

## Riesgo en tiempo discreto

Para  $T$  discreta con soporte  $\{u_1, u_2, \dots\}$ :

$$h(t) = P(T = t \mid T \geq t)$$
$$h_k = \frac{P(T = u_k)}{P(T \geq u_k)} = \frac{f(u_k)}{S(u_{k-1})}$$

Usando  $f(u_k) = S(u_{k-1}) - S(u_k)$ , se obtiene:

$$h_k = 1 - \frac{S(u_k)}{S(u_{k-1})}$$

## Relaciones discretas clave

Función de supervivencia:

$$S(t) = \prod_{u_k \leq t} (1 - h_k)$$

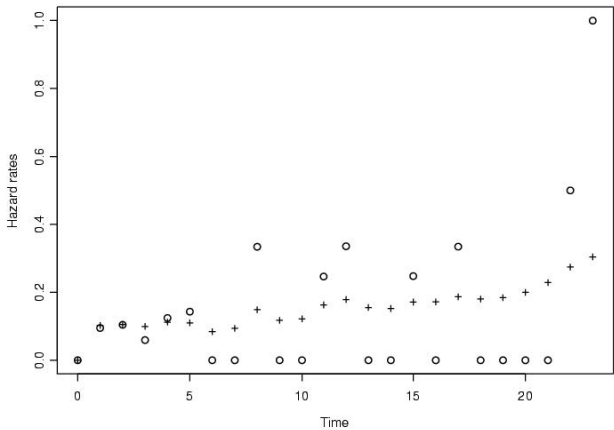
Función de densidad:

$$f(u_j) = h_j \prod_{k < j} (1 - h_k)$$

En demografía,  $h(t)$  representa la probabilidad de morir en el momento  $t$  dado que se ha sobrevivido hasta  $t$ .

## Ejemplos de riesgo discreto

CIMAT



## Riesgo en tiempo continuo

CIMAT

$$h(t) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P(t < T \leq t + \varepsilon \mid T \geq t) = \frac{f(t)}{S(t)}$$

Como  $F(t) = 1 - S(t)$ , entonces:

$$h(t) = -\frac{d}{dt} \log S(t)$$

Al integrar:

$$\log S(t) = -\int_0^t h(u) du$$

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

$h(t)\varepsilon$  es la probabilidad **aproximada** de que un evento ocurra en el siguiente instante dado que el individuo ha sobrevivido hasta  $t$ .

## Visualización de funciones

CIMAT

36

## Riesgo acumulado discreto

CIMAT

1. Suma directa:

$$H(t) = \sum_{t_k \leq t} h_k$$

2. Log-transformación:

$$H(t) = -\sum_{t_k \leq t} \log(1 - h_k)$$

Ambas son **monótonas no decrecientes**.

38

## Riesgo acumulado continuo

CIMAT

$$H(t) = \int_0^t h(u) du \quad \Rightarrow \quad S(t) = \exp\{-H(t)\}$$

Si  $S(\infty) = 0$ , entonces  $H(\infty) = \infty$ .

40



CIMAT

Hazard Rates, Survival Functions, Probability Density Functions, and Expected Lifetimes for Some Common Parametric Distributions

Distribution	Hazard Rate $h(x)$	Survival Function $S(x)$	Probability Density Function $f(x)$	Mean $E(X)$
Exponential $\lambda > 0, x \geq 0$	$\lambda$	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0, x \geq 0$	$\alpha \lambda x^{\alpha-1}$	$\exp[-\lambda x^\alpha]$	$\alpha \lambda x^{\alpha-1} \exp(-\lambda x^\alpha)$	$\frac{\Gamma(1 + 1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x; \beta)^*$	$\frac{\lambda^\beta x^{\beta-1} \exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^2)$
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha x^{\alpha-1} \lambda}{1 + \lambda x^\alpha}$	$\frac{1}{1 + \lambda x^\alpha}$	$\frac{\alpha x^{\alpha-1} \lambda}{[1 + \lambda x^\alpha]^2}$	$\frac{\pi \csc(\pi/\alpha)}{\alpha \lambda^{1/\alpha}}$ if $\alpha > 1$



## ¿Qué son la censura y el truncamiento?

En análisis de supervivencia, no siempre se observa el tiempo completo hasta que ocurre un evento.

- **Censura:** el evento de interés no se ha observado completamente.
- **Truncamiento:** ciertas observaciones no entran en el estudio porque el evento ocurre fuera del intervalo observable.

Estas condiciones afectan la estimación de funciones como  $S(t)$  y  $h(t)$ , y requieren técnicas específicas.

# Parte 5: Censura y truncamiento



## Censura y truncamiento

### ¿Qué es la censura?

- La **censura** ocurre cuando **no se observa completamente** el tiempo de fallo.
- Es común en estudios longitudinales, donde algunos individuos:
  - **No han fallado** al final del estudio,
  - **Ingresan tarde** al seguimiento,
  - **O se pierde el seguimiento.**

#### Tipos de censura:

- **Por la derecha:** solo sabemos que el evento ocurrió después de cierto tiempo.
- **Por la izquierda:** solo sabemos que ocurrió antes de cierto tiempo.
- **Por intervalo:** solo sabemos que ocurrió entre dos tiempos.

### ¿Qué es el truncamiento?

- El **truncamiento** ocurre cuando **ciertas observaciones nunca se registran** debido al diseño del estudio.

#### Ejemplos:

- **Truncamiento por la izquierda:** sólo se incluyen individuos cuyo evento ocurre **después** de cierto punto.
- **Truncamiento por la derecha:** se excluyen individuos cuyo evento ocurre **después** de cierto punto.

#### Implicación:

- Afecta **quién entra al estudio** (selección), no solo cómo se mide el tiempo.



### Notas para el instructor:

Puedes acompañar esta diapositiva con una **gráfica tipo línea de tiempo** que contraste **censura vs truncamiento** (e.g. tipo Klein & Moeschberger).

- Si se quiere profundizar, añade en otra diapositiva cómo **afectan a la función de verosimilitud**.
- Si el público es nuevo en supervivencia, considera ilustrar con ejemplos cotidianos:
  - Censura: "no sabemos cuándo murió un paciente, solo que estaba vivo al final del estudio".
  - Truncamiento: "personas que murieron antes de entrar al estudio no están en el conjunto de datos".



## Tipos de censura por la derecha

### ¿Cómo se genera la censura?

En estudios de supervivencia, es común que **no se observe completamente** el tiempo de falla. Esto ocurre mediante distintos mecanismos:

- **Censura tipo I**
- **Censura tipo II**
- **Censura aleatoria**



## Censura tipo I

**Definición:**  
Se observa el tiempo de supervivencia  $T_i$ ; solo si ocurre antes de un tiempo de **censura predeterminado**  $C_i$ .  
Si  $T_i > C_i$ , entonces el dato está censurado.

#### Notación formal:

- Observamos el par  $(t_i, \delta_i)$ , donde
  - $t_i = \min(T_i, C_i)$
  - $\delta_i = I(T_i \leq C_i)$
- Si  $\delta_i = 1$ : observación completa
- Si  $\delta_i = 0$ : censura por la derecha

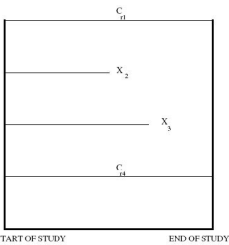


Figure 3.1 Example of Type I censoring



## Ejemplo de censura tipo I

CIMAT En un estudio toxicológico, ratones reciben un carcinógeno. Se observa su supervivencia hasta cierto tiempo límite. Los ratones aún vivos en ese punto son sacrificados (censurados).

Importante: Puede haber **múltiples tiempos de censura**, dependiendo del diseño experimental.

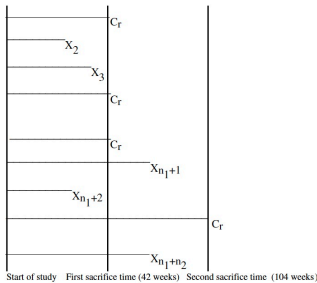


Figure 3.2 Type I censoring with two different sacrifice times

## Censura tipo II

CIMAT **Definición:** El estudio se **detiene al observar la falla de los primeros  $r < n$  sujetos**.

- Se observan los tiempos  $T_{(1)}, T_{(2)}, \dots, T_{(r)}$
- Los  $n - r$  sujetos restantes están censurados.

Notación:

- Tiempo de censura común:  $C = T_{(r)}$
- Censura si  $T_i > C$

Aplicación típica: pruebas de resistencia de equipos que se detienen al fallar cierto número de unidades.

Ejemplo: Se prueban 20 motores, pero se termina el estudio tras la falla de los primeros 10.

## Censura tipo I generalizada

CIMAT Cada individuo entra en un momento distinto al estudio, pero el final del estudio está predeterminado.

- Cada sujeto tiene su propio tiempo de censura **fijo**.
- Este diseño genera **censura tipo I generalizada**.

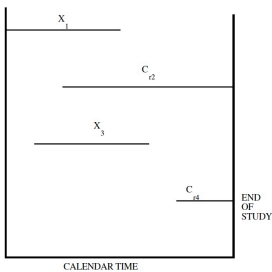


Figure 3.3 Generalized Type I censoring when each individual has a different starting time

## Censura aleatoria

CIMAT **Definición:** El tiempo de censura  $C_i$  es una **variable aleatoria**, diferente para cada individuo.

Ejemplos comunes:

- Salida del estudio
- Muerte por otra causa
- Migración o pérdida de contacto
- Hospital deja de aceptar al paciente

Tipos de censura aleatoria

- No informativa:**  $C_i \perp T_i$   
→ tratable como censura tipo I
- Informativa:**  $C_i$  depende de  $T_i$   
→ requiere modelos avanzados

Ejemplo aplicado: En estudios con pacientes de diálisis, el evento de interés puede ser fallas por infección, pero se censura por muerte o salida del hospital.

## Conclusión

- La **censura por la derecha** se clasifica según su origen: planificado, condicional o aleatorio.
- Es fundamental para construir correctamente la **función de verosimilitud**.
- Un mal manejo de la censura puede sesgar estimaciones.

## Censura por la izquierda e intervalo

CIMAT **¿Qué es la censura por la izquierda?**

- Ocorre cuando el evento de interés sucede antes de un tiempo de observación conocido.
- Es decir, sabemos que el evento ya ocurrió, pero no cuándo exactamente.

Definición formal: Sea  $C_i$  el tiempo de censura por la izquierda y  $T_i$  el tiempo de falla.

- Si  $T_i \geq C_i$ : observación completa.
- Si  $T_i < C_i$ : **censura por la izquierda**.

Notación:

$$t_i = \max(T_i, C_i), \quad \delta_i = I(T_i \geq C_i)$$

Ejemplo 1

Adolescente declara: "Si consumí marihuana, pero no recuerdo cuándo".  
→ El evento ocurrió antes de su edad actual, pero se desconoce el momento exacto.

Ejemplo 2

Un niño ya sabe realizar una tarea cuando entra al estudio.  
→ El aprendizaje ocurrió antes de la observación inicial.

## Censura doble (izquierda y derecha)

**Definición:**  
Una observación está **doblemente censurada** si se desconoce si el evento ocurrió antes o después de un cierto rango.

- Combina censura por la izquierda y la derecha.
- Común en estudios transversales o con límites temporales de observación.

**Notación generalizada:**

$$t_i = \max\{\min(T_i, C_r), C_l\}, \quad \delta_i = \begin{cases} 1, & \text{tiempo exacto} \\ 0, & \text{censura por la derecha} \\ -1, & \text{censura por la izquierda} \end{cases}$$

**Ejemplo 1 – Marihuana**

"Nunca la he usado" → censura por la derecha  
"Si la usé pero no recuerdo cuándo" → izquierda  
"La usé a los 15" → observación completa

**Ejemplo 2 – Aprendizaje infantil**

Algunos niños no aprenden durante el estudio → censura por la derecha  
Otros ya sabían antes de iniciar → censura por la izquierda

## Recomendaciones

**Identificar con claridad el momento de entrada al estudio y el horizonte de observación.**

- Siempre registrar **si se trata de censura por la izquierda, derecha o ambas**.
- Verificar si la censura es **informativa o no informativa**.

## Ejemplo 1 – Estudio del Corazón de Framingham

**En este estudio longitudinal, los eventos de enfermedad coronaria (CHD) pueden registrarse con precisión.**

**Sin embargo:**

- La aparición de **angina de pecho** se detecta solo **entre dos visitas clínicas**, con varios años de diferencia.

→ El tiempo exacto es desconocido, pero **ocurrió dentro del intervalo** entre exámenes.

## Comparación gráfica sugerida

**Agregar una figura como la siguiente para reforzar visualmente)**

Tipo de observación	Línea de tiempo
Exacta	Evento ocurre entre observación inicial y final
Censura por la derecha	Línea que termina sin evento registrado
Censura por la izquierda	Línea que empieza con evento ya ocurrido
Doble censura	Solo se sabe que el evento ocurrió fuera del intervalo de observación

## Censura por intervalo

**¿Qué es la censura por intervalo?**

- La **censura por intervalo** ocurre cuando **el evento sucede entre dos visitas clínicas**, pero **no se conoce el momento exacto**.

**Interpretación:**

Se sabe que el sujeto **no había fallado antes del tiempo  $L_i$** , pero **si lo ha hecho antes o en el tiempo  $R_i$** .

**Notación formal:**

$$L_i < T_i \leq R_i$$

**Donde:**

- $L_i$  = última vez que se observó sin evento
- $R_i$  = primera vez que se detecta el evento

Puedes pensar esta censura como una observación **con ventana de tiempo**, en la que el evento ocurre **dentro de un intervalo** que puede variar por sujeto.

**Posibles causas:**

- Visitas clínicas programadas
- Limitaciones de seguimiento continuo

## Ejemplo 2 – Estudio de radioterapia

**Se estudió el efecto cosmético en mujeres con cáncer de mama tras radioterapia (con o sin quimioterapia).**

- Se realizaron controles **cada 4 a 6 meses**, luego más espaciados.
- El evento de interés: **retracción severa del seno**.
- Solo se sabía si **ocurrió entre dos visitas**, o si **nunca se observó** (censura por la derecha).

→ Algunas pacientes presentaron **censura por intervalo**, y otras, **por la derecha**.

Consideraciones finales

CIMA

Censura por intervalo requiere un tratamiento especial en modelos de supervivencia:

No puede tratarse como observación exacta ni como censura por la derecha.

Métodos como EM, modelos de verosimilitud para intervalos, o análisis de datos agrupados son apropiados.

Truncamiento

CIMA

¿Qué es el truncamiento?

El truncamiento ocurre cuando ciertos individuos no aparecen en el estudio, porque su tiempo de falla está fuera de una ventana de observación.

Diferencia clave con la censura:

Censura → se observa parcialmente

Truncamiento → no se observa en absoluto

Definición formal:

Se observa  $T_i$  solo si  $T_i \in (U_i, V_i)$

Imaginemos una ventana de observación:  
Si el evento ocurre antes de entrar a la ventana o después de que cierra, el sujeto no entra al estudio.  
Esto es truncamiento, no censura.

Truncamiento por la izquierda

CIMA

Definición:  
Solo se observan individuos cuyo tiempo de evento supera un umbral inferior:

$$T_i > U_i$$

También conocido como **entrada retardada**: el sujeto **debió sobrevivir** lo suficiente para entrar al estudio.

Ejemplo – Centro de retiro

En Channing House, solo se estudian residentes que **lograron ingresar**.  
Quienes murieron **antes** de tener edad para ingresar, **no aparecen** en el estudio.  
→ **Truncamiento por la izquierda**

Truncamiento por la derecha

CIMA

Definición:  
Solo se incluyen sujetos cuyo evento ocurre **antes de un umbral superior**:

$$T_i < V_i$$

Esto puede ocurrir en **estudios retrospectivos con fecha de corte**.

Ejemplo – Estudio del SIDA

Solo se incluyen pacientes que desarrollaron SIDA **antes del 30 de junio de 1986**.  
Aquellos cuya enfermedad apareció después, **no fueron observados**.  
→ **Truncamiento por la derecha**

Truncamiento y censura combinados

CIMA

Es **común** que los estudios de supervivencia combinen:

Truncamiento por la izquierda (entrada tardía)

Censura por la derecha (seguimiento limitado)

Ejemplo típico:

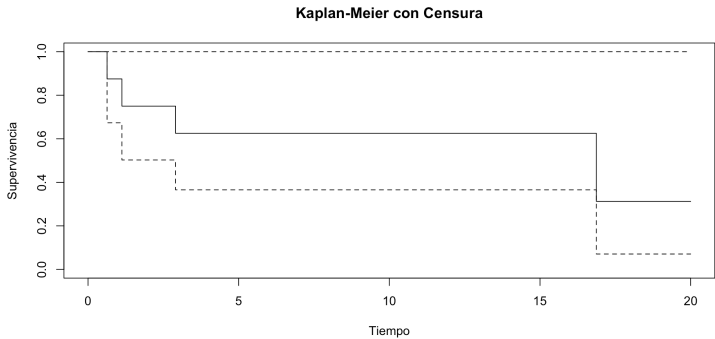
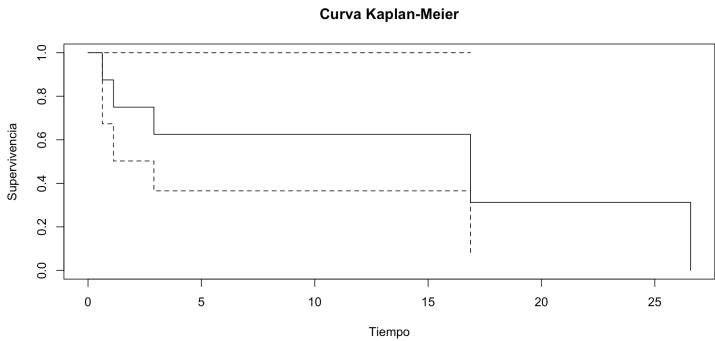
Un paciente entra al estudio tras cumplir ciertos criterios (truncamiento), pero el estudio termina antes de que fallezca (censura).

Visualización sugerida

CIMA

Agregar figura con líneas de tiempo para comparar:)

Tipo de situación	Línea de tiempo
Truncamiento por izquierda	Evento ocurre antes de entrada: no se observa
Truncamiento por derecha	Evento ocurre después de fecha límite: no se observa
Censura por derecha	Evento ocurre después de seguimiento: sí se observa parcialmente
Observación completa	Evento ocurre dentro del periodo: sí se observa



```
Call:
survreg(formula = Surv(time, event) ~ 1, data = data_sim, dist = "exponential")
      Value Std. Error      z      p
(Intercept) 2.671      0.447  5.97 2.3e-09

Scale fixed at 1

Exponential distribution
Loglik(model)= -18.4   Loglik(intercept only)= -18.4
Number of Newton-Raphson Iterations: 5
n= 8

Call:
survreg(formula = Surv(time, event) ~ 1, data = data_sim, dist = "weibull")
      Value Std. Error      z      p
(Intercept) 2.686      0.525  5.12 3.1e-07
Log(scale)  0.154      0.367  0.42  0.68

Scale= 1.17

Weibull distribution
Loglik(model)= -18.3   Loglik(intercept only)= -18.3
Number of Newton-Raphson Iterations: 6
n= 8
```

- Censura por la derecha, izquierda e intervalos
- Truncamiento (izquierdo/ derecho)
- Impacto en el análisis de supervivencia

- Distribuciones exponencial, Weibull, log-logístico
- Ajuste de modelos en R

## La función de distribución acumulada empírica (FDAE)

Dada una muestra de tiempos de falla sin censura:

$$\hat{F}(t) = \frac{\#\{T_i \leq t\}}{n}$$

Es un estimador escalonado, que da saltos en cada observación.  
La función de supervivencia empírica se define como:

$$\hat{S}(t) = 1 - \hat{F}(t)$$

**Limitación:** no puede manejar adecuadamente datos censurados.

## Estimador de Kaplan-Meier

Quando hay censura, la FDAE no es válida. Kaplan-Meier estima la función de supervivencia como:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

donde:

- $d_i$ : número de eventos en el tiempo  $t_i$
- $n_i$ : número de individuos en riesgo justo antes de  $t_i$

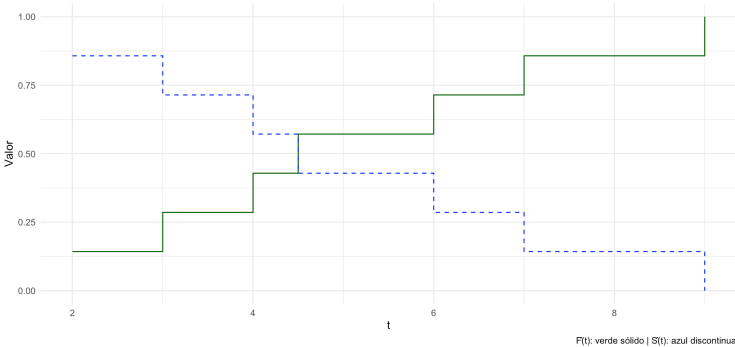
Es un estimador escalonado que **ajusta el denominador** cuando hay censura.

## Comparación conceptual

Característica	FDAE	Kaplan-Meier
Usa solo eventos	✓	✓
Maneja censura	✗	✓
Escalonada	✓	✓
Basada en conteos simples	✓	✗ (ajusta denominadores)

## Ejemplo en R: FDAE

FDAE y supervivencia empírica sin censura

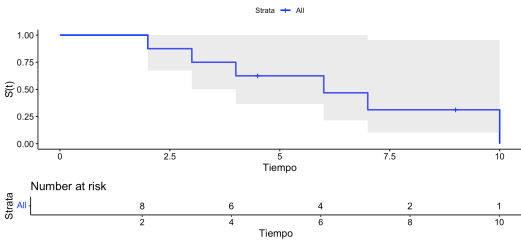


## Ejemplo en R: Kaplan-Meier

Tabla de tiempos y estatus de censura

ID	tiempo	evento
Ind 1	2.0	1
Ind 2	3.0	1
Ind 3	4.0	1
Ind 4	4.5	0
Ind 5	6.0	1
Ind 6	7.0	1
Ind 7	9.0	0
Ind 8	10.0	1

Estimación de Kaplan-Meier



Call: survfit(formula = surv\_obj ~ 1, data = datos)

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2	8	1	0.875	0.117	0.673	1.000
3	7	1	0.750	0.153	0.503	1.000
4	6	1	0.625	0.171	0.365	1.000
6	4	1	0.469	0.187	0.215	1.000
7	3	1	0.312	0.178	0.102	0.955
10	1	1	0.000	NaN	NA	NA