

# Análisis de Supervivencia

## Introducción

Sergio M. Nava Muñoz

2025-06-01

## 1 Introducción al Análisis de Supervivencia

### 1.1 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

### 1.2 ¿Qué es el análisis de supervivencia?

#### Análisis de Supervivencia

El análisis de supervivencia, también conocido como análisis de **tiempo hasta evento**, es un conjunto de técnicas estadísticas diseñadas para manejar datos censurados y modelar la distribución del tiempo hasta la ocurrencia de un evento de interés, ver Klein & Moeschberger (2003).

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
- 

## 1.3 ¿Dónde se aplica el análisis de supervivencia?

### 1.3.1 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.

---

## 1.4 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

---

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

Para interpretar adecuadamente los datos, es necesario:

### 1.5.1 1. Evento de origen

→ ¿Desde cuándo empieza a contarse el tiempo?

### 1.5.2 2. Escala de medición

→ ¿Cómo se mide el tiempo? (reloj, kilómetros, ciclos)

### 1.5.3 3. Evento de fin

→ ¿Qué se considera “fallo”, “recuperación”, o “evento”?

## 2 Ejemplos

### 2.1 Ejemplos varios

**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

## 2.2 Ejemplos de datos de supervivencia en bioestadística

A continuación se presentan algunos ejemplos de datos de supervivencia. Estos ejemplos fueron obtenidos de Klein & Moeschberger (2003).

### 2.2.1 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 compara 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 prednisone. 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 <sup>+</sup>
4	Complete Remission	12	23
5	Complete Remission	8	22
6	Partial Remission	17	6
7	Complete Remission	2	16
8	Complete Remission	11	34 <sup>+</sup>
9	Complete Remission	8	32 <sup>+</sup>
10	Complete Remission	12	25 <sup>+</sup>
11	Complete Remission	2	11 <sup>+</sup>
12	Partial Remission	5	20 <sup>+</sup>
13	Complete Remission	4	19 <sup>+</sup>
14	Complete Remission	15	6
15	Complete Remission	8	17 <sup>+</sup>
16	Partial Remission	23	35 <sup>+</sup>
17	Partial Remission	5	6
18	Complete Remission	11	13
19	Complete Remission	4	9 <sup>+</sup>
20	Complete Remission	1	6 <sup>+</sup>
21	Complete Remission	8	10 <sup>+</sup>

<sup>+</sup> Censored observation

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

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 medula 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.

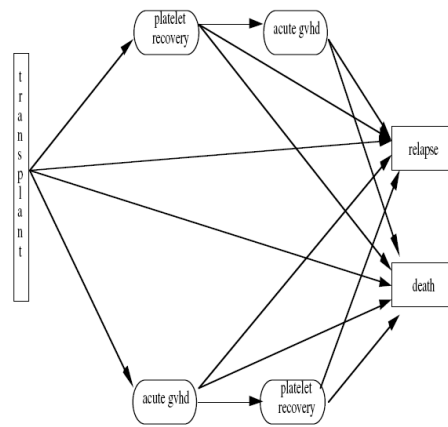


Figure 1.1 Recovery Process from a Bone Marrow Transplant

---

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

---

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

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.

TABLE D.1  
(continued)

$g$	$T_1$	$T_2$	$\delta_1$	$\delta_2$	$\delta_3$	$T_A$	$\delta_A$	$T_C$	$\delta_C$	$T_P$	$\delta_P$	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$	$Z_6$	$Z_7$	$Z_8$	$Z_9$	$Z_{10}$
1	2081	2081	0	0	0	67	1	121	1	13	1	26	33	1	0	1	1	98	0	1	0
1	1602	1602	0	0	0	1602	0	139	1	18	1	21	37	1	1	0	0	1720	0	1	0
1	1496	1496	0	0	0	1496	0	307	1	12	1	26	35	1	1	1	0	127	0	1	0
1	1462	1462	0	0	0	70	1	95	1	13	1	17	21	0	1	0	0	168	0	1	0
1	1433	1433	0	0	0	1433	0	236	1	12	1	32	36	1	1	1	1	93	0	1	0
1	1377	1377	0	0	0	1377	0	123	1	12	1	22	31	1	1	1	1	2187	0	1	0
1	1330	1330	0	0	0	1330	0	96	1	17	1	20	17	1	0	1	1	1006	0	1	0
1	996	996	0	0	0	72	1	121	1	12	1	22	24	1	0	0	0	1319	0	1	0
1	226	226	0	0	0	226	0	226	0	10	1	18	21	0	1	0	0	208	0	1	0
1	1199	1199	0	0	0	1199	0	91	1	29	1	24	40	1	1	0	1	174	0	3	1
1	1111	1111	0	0	0	1111	0	1111	0	22	1	19	28	1	1	0	1	236	0	3	1
1	530	530	0	0	0	38	1	84	1	34	1	17	28	1	1	0	0	151	0	3	1

TABLE D.1

Data on 137 Bone Marrow Transplant Patients

$g$ —Disease group

1-ALL

2-AML low-risk

3-AML high-risk

$T_1$ —Time (in days) to death or on study time

$T_2$ —Disease-Free survival time (time to relapse, death or end of study)

$\delta_1$ —Death indicator

1-Dead 0-Alive

$\delta_2$ —Relapse indicator

1-Relapsed 0-Disease-Free

$\delta_3$ —Disease-Free survival indicator

1-Dead or relapsed 0-Alive disease-free

$T_A$ —Time (in days) to acute graft-versus-host disease

$\delta_A$ —Acute graft-versus-host disease indicator

1-Developed acute graft-versus-host disease

0-Never developed acute graft-versus-host disease

$T_C$ —Time (in days) to chronic graft-versus-host disease

$\delta_C$ —Chronic graft-versus-host disease indicator

1-Developed chronic graft-versus-host disease

0-Never developed chronic graft-versus-host disease

$T_P$ —Time (in days) to return of platelets to normal levels

$\delta_P$ —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

$Z_8$ —FAB

1-FAB Grade 4 Or 5 and AML 0-Otherwise

$Z_9$ —Hospital

1-The Ohio State University 2-Alfred

3-St. Vincent 4-Hahnemann

$Z_{10}$ —MTX used as a graft-versus-host-prophylactic

1-Yes 0-No

```

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      885      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

```

---

### 2.2.5 Ejemplo Tiempo al primer uso de marihuana.

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*

<i>Age</i>	<i>Number of Exact Observations</i>	<i>Number Who Have Yet to Smoke Marijuana</i>	<i>Number Who Have Started Smoking at an Earlier Age</i>
10	4	0	0
11	12	0	0
12	19	2	0
13	24	15	1
14	20	24	2
15	13	18	3
16	3	14	2
17	1	6	3
18	0	0	1
>18	4	0	0

---

## 2.2.6 Tiempo a desarrollar sida.

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 en 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.

## 2.2.7 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
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



Infection Time	Adult Induction Time	Child Induction Time
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

## 2.3 Ejemplo de datos simulados $(t_i, \delta_i)$

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

Table 2: 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

## 2.4 Actividad en clase: Diseño rápido de un ejemplo de datos de supervivencia

### 2.4.1 Objetivo de la actividad

Durante esta actividad breve, los estudiantes identificarán los elementos esenciales de un conjunto de datos que pueda ser analizado mediante técnicas de análisis de supervivencia.

### 2.4.2 Duración sugerida

15 minutos

### 2.4.3 Modalidad

- Individual
  - [Hoja de Trabajo](#)
- 

### 2.4.4 Instrucciones

1. **Piensa en una situación real o simulada** donde se puedan observar tiempos hasta que ocurra un evento específico (por ejemplo: tiempo hasta el abandono escolar, tiempo hasta la reincidencia en consumo, duración de un aparato, etc.).
  2. **Completa los siguientes elementos clave de tu ejemplo:**
    - **Contexto:** ¿en qué situación o ámbito ocurre?
    - **Evento de origen (inicio del seguimiento):** ¿cuándo empieza a contar el tiempo?
    - **Escala de tiempo:** ¿en qué unidades se mide el tiempo (días, semanas, meses...)?
    - **Evento de fin (evento de interés):** ¿qué suceso representa el “evento”?
    - **¿Hay censura posible?** ¿cuándo podría ocurrir?
    - *(Opcional)* **Covariables:** ¿qué otras características podrían influir en el tiempo al evento?
  3. Escribir en el espacio compartido [Hoja de Trabajo](#).
- 

### 2.4.5 Ejemplo rápido

- **Contexto:** Estudio sobre empleados de una empresa.
- **Evento de origen:** Fecha de contratación.
- **Escala de tiempo:** Meses desde el ingreso.
- **Evento de fin:** Renuncia o despido.

- **Censura:** Si el empleado sigue activo al final del estudio.
- **Covariables:** Edad, departamento, tipo de contrato.

### 3 Tipos de datos: eventos, censura, truncamiento

#### 3.1 Censura y truncamiento

##### 3.1.1 ¿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.

##### 3.1.2 ¿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.

## 3.2 Tipos de censura por la derecha

### 3.2.1 ¿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
- 

## 3.3 Censura tipo I

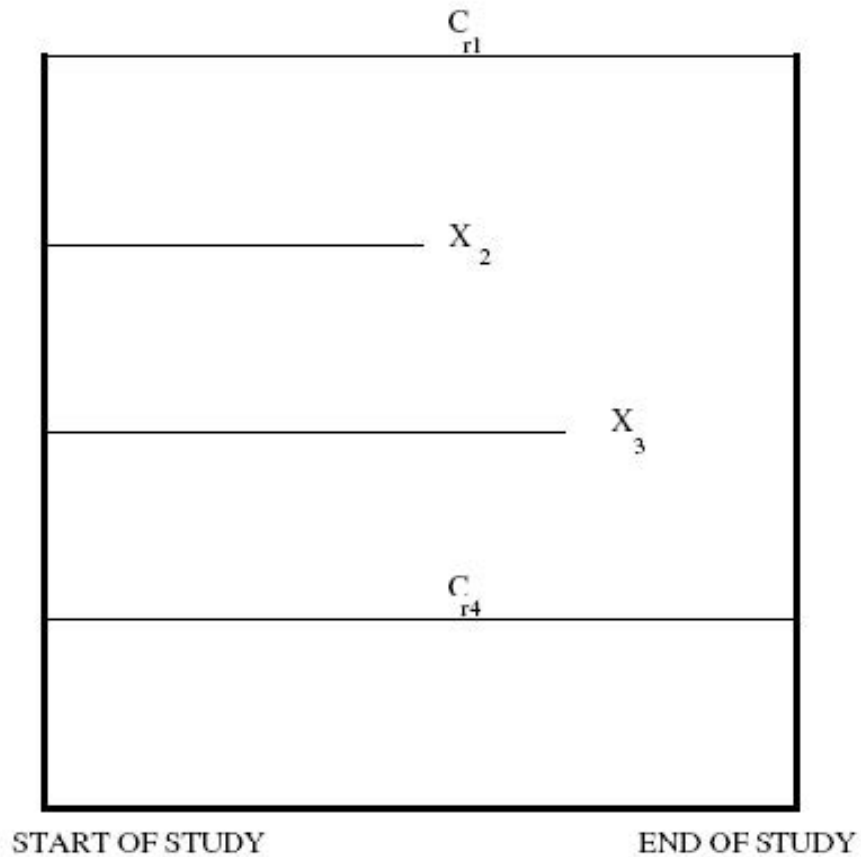
### Definición:

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

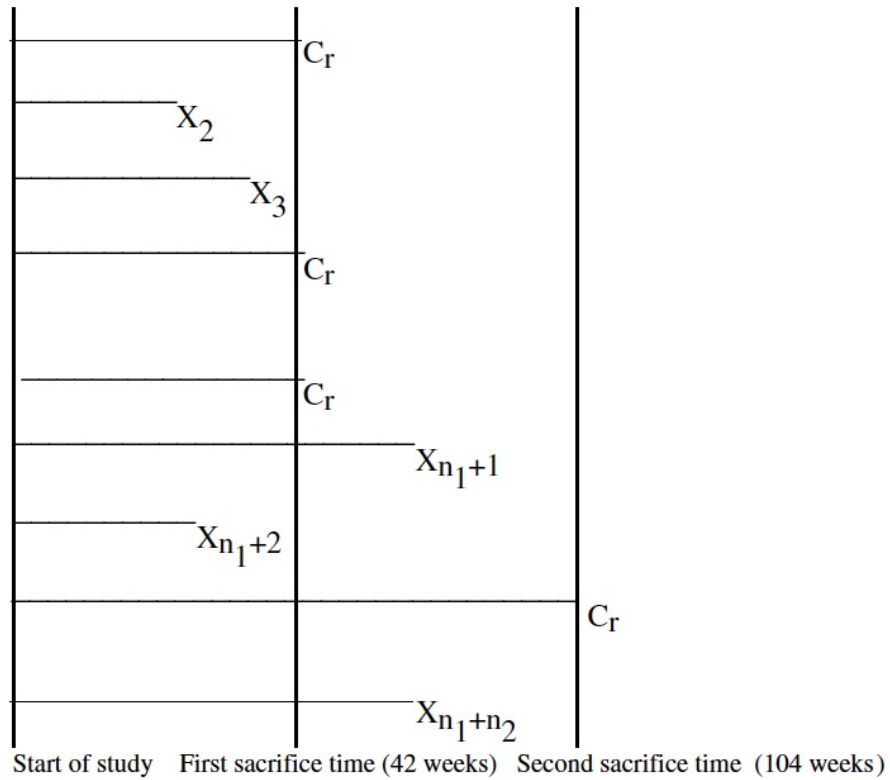


**Figure 3.1** *Example of Type I censoring*

### 3.3.1 Ejemplo de censura tipo I

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.

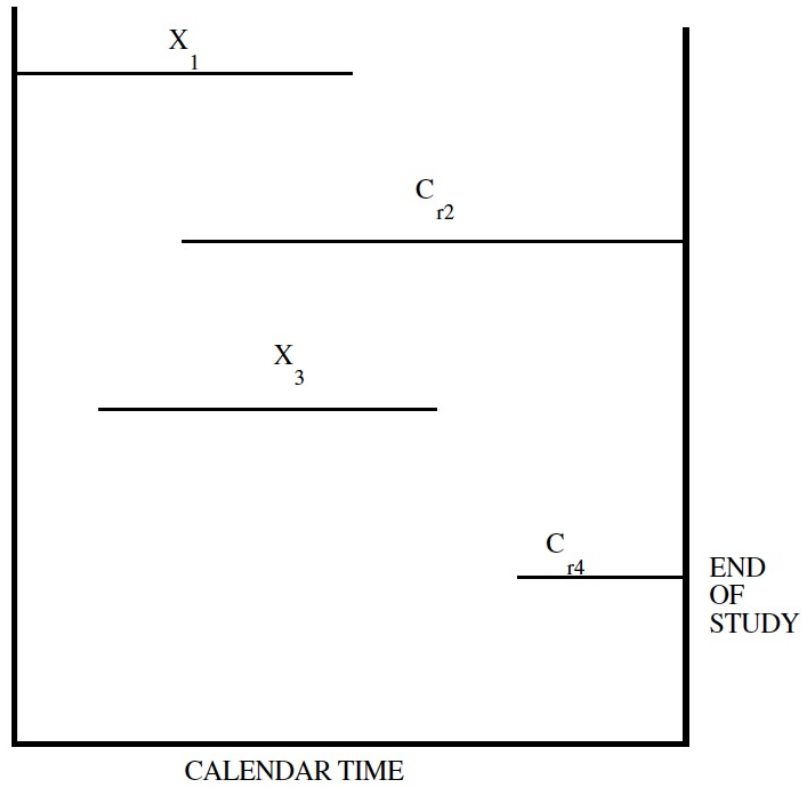


**Figure 3.2** *Type I censoring with two different sacrifice times*

### 3.3.2 Censura tipo I generalizada

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*

### 3.4 Censura tipo II

**Definición:**

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

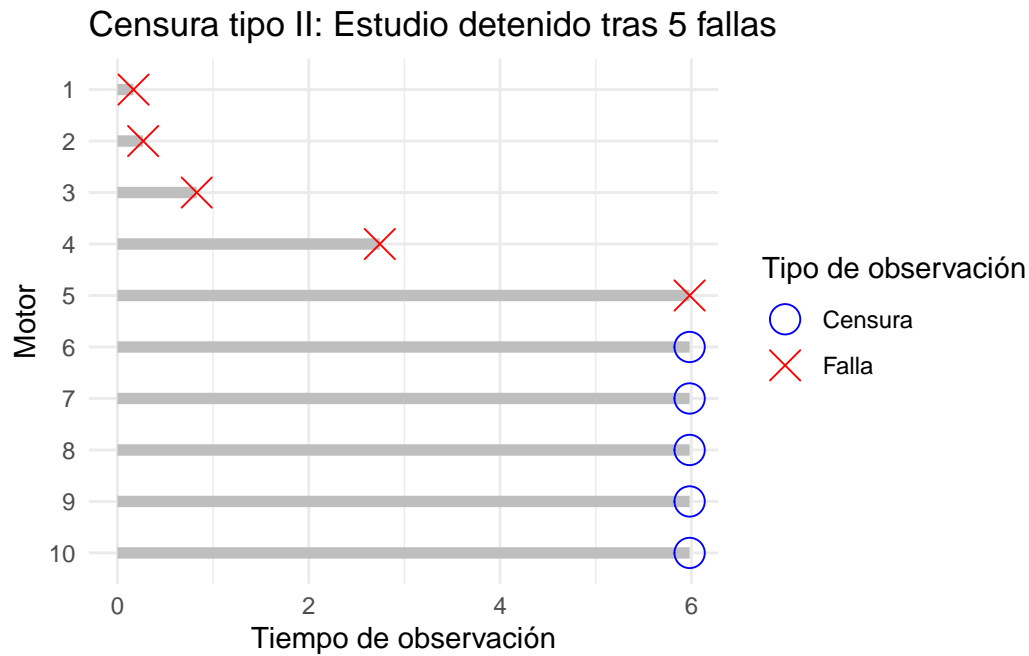
- Se observan los tiempos  $T_{(1)}, T_{(2)}, \dots, T_{(r)}$
- Los  $n - r$  sujetos restantes están censurados.

**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 10 motores, pero se termina el estudio tras la falla de los primeros 5.



### 3.5 Censura aleatoria

**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

*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.



### 3.5.1 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
- 

## 3.6 Censura por la izquierda e intervalo

### 3.6.1 ¿Qué es la censura por la izquierda?

- Ocurre 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  $C_l$  el tiempo de censura por la izquierda y  $T_i$  el tiempo de falla.

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

#### Notación:

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

### Ejemplo 1

Adolescente declara:

*“Sí 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.

---

### 3.7 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
- “Sí la usé pero no recuerdo cuándo” → izquierda
- “La usé a los 15” → observación completa

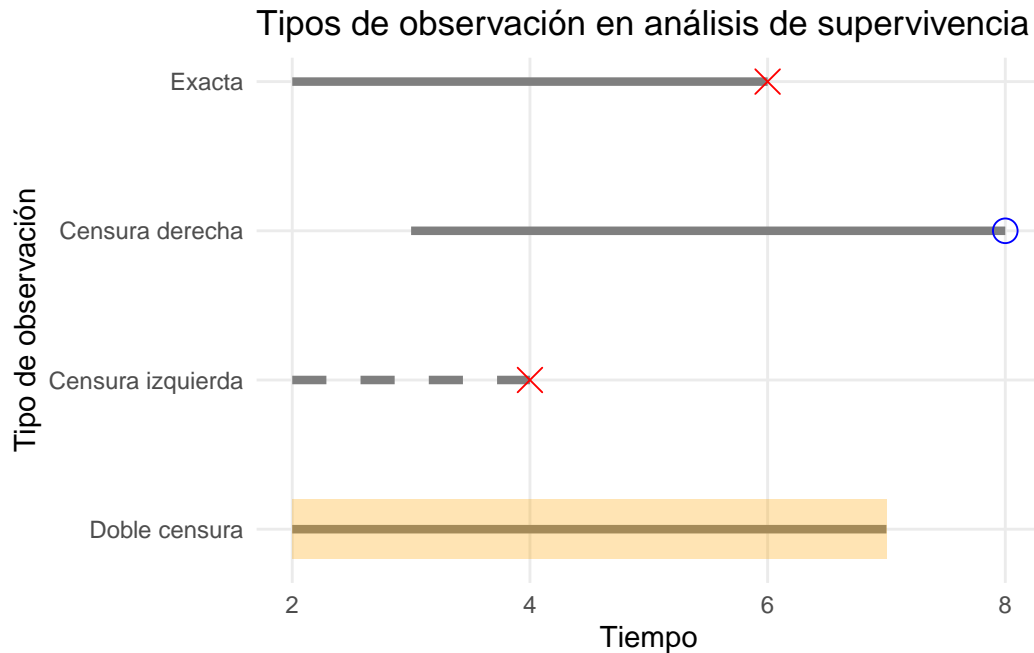
#### Ejemplo 2 – Aprendizaje infantil

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

---

### 3.8 Comparación

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



### 3.9 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**.

### 3.10 Censura por intervalo

#### 3.10.1 ¿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 sí 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

### 3.10.2 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.

### 3.10.3 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**.

---

### 3.11 Actividad: Identificación y análisis de datos de supervivencia

**3.11.0.1 Objetivo** Aplicar los conceptos fundamentales del análisis de supervivencia identificando un conjunto de datos relevante, describiendo su estructura temporal, y evaluando la presencia y tipo de censura.

#### 3.11.0.2 Instrucciones

1. **Buscar o seleccionar un conjunto de datos** que permita aplicar análisis de supervivencia. Puede ser:
  - Un conjunto de datos **propio** (proyecto, tesis, trabajo profesional).
  - Un conjunto de datos **público** (Kaggle, UCI, CRAN, etc.).
2. **Describir el contexto del estudio**, incluyendo:
  - La **unidad de análisis** (por ejemplo: paciente, máquina, usuario).
  - La **variable de tiempo** (por ejemplo: días hasta falla, semanas hasta abandono).
  - El **evento de interés** (por ejemplo: muerte, falla, compra, abandono).
3. **Explicar la estructura temporal del estudio**, atendiendo a las siguientes recomendaciones:
  - Identificar con claridad el **momento de entrada** al estudio y el **horizonte de observación**.
  - Registrar si se trata de **censura por la izquierda, derecha o ambas**.
  - Verificar si la censura es **informativa o no informativa**.
4. **Entregar un breve reporte (1-2 cuartillas)** que contenga:
  - Descripción del conjunto de datos.
  - Contexto del estudio y definición del evento.
  - Discusión sobre tipo(s) de censura y su naturaleza.
  - Una tabla ilustrativa con al menos 10 observaciones con las columnas: ID, tiempo, status (evento = 1, censura = 0).

## 3.12 Truncamiento

### 3.12.1 ¿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.

---

## 3.13 Truncamiento por la izquierda

**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**

---

### 3.14 Truncamiento por la derecha

#### 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**

---

### 3.15 Truncamiento y censura combinados

- 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).

## 4 Visualización

### 4.1 Ejemplo: Ensayo clínico con cáncer

paciente	entrada	fin	evento
1	2000	2007	0
2	2000	2006	1
3	2001	2007	0
4	2002	2007	0
5	2002	2004	1
6	2002	2006	1

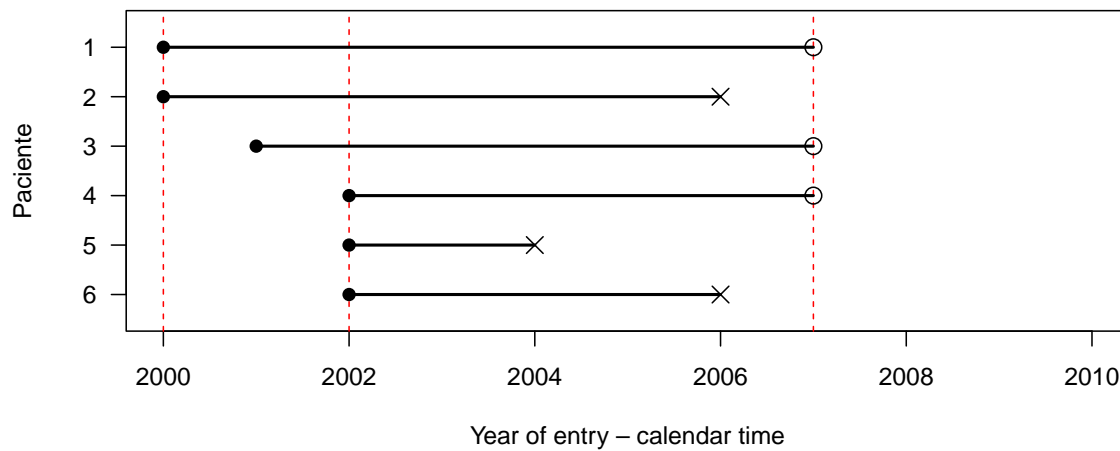


Figure 1: Reclutamiento y seguimiento

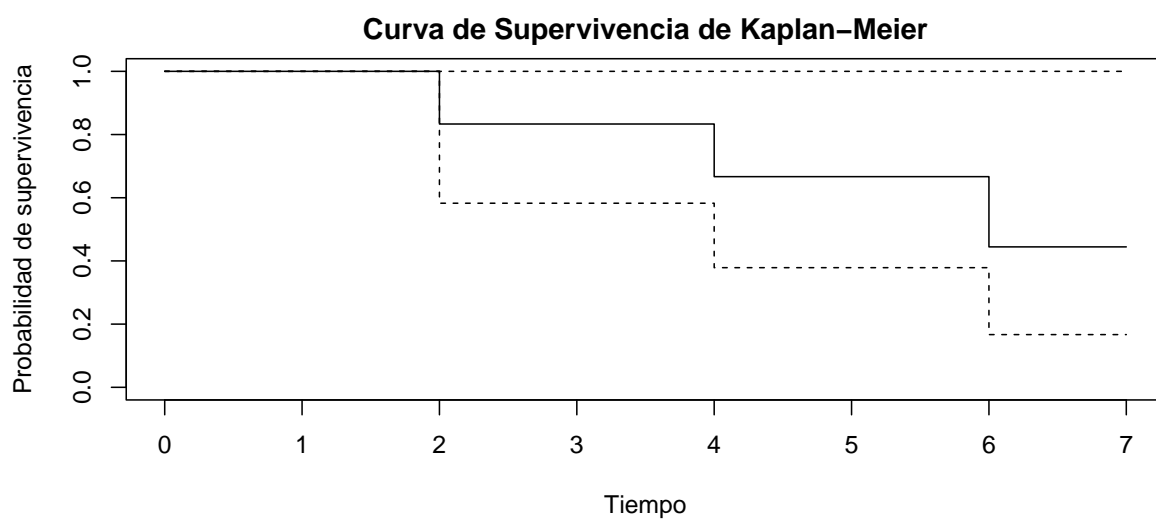
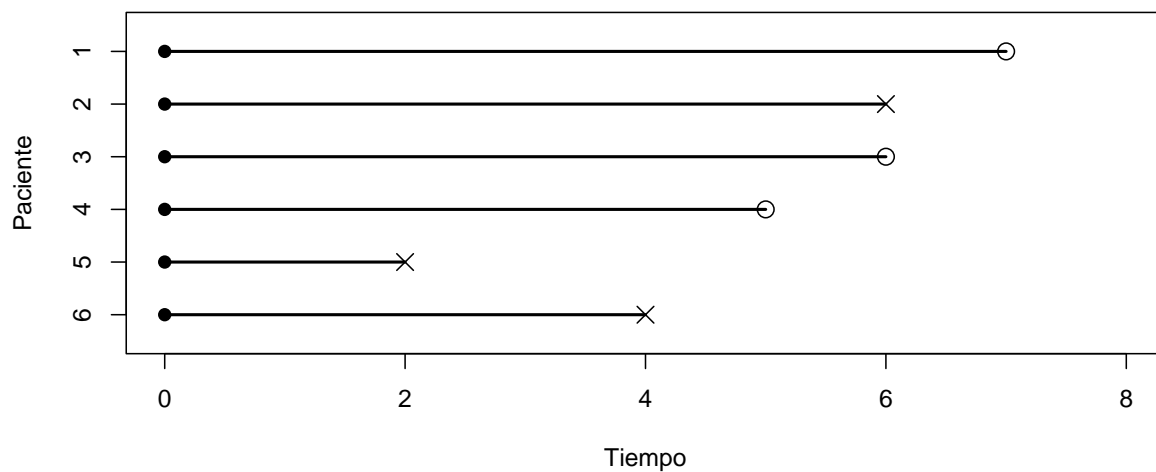
## 4.2 Representación gráfica del seguimiento

Table 5: Ejemplo

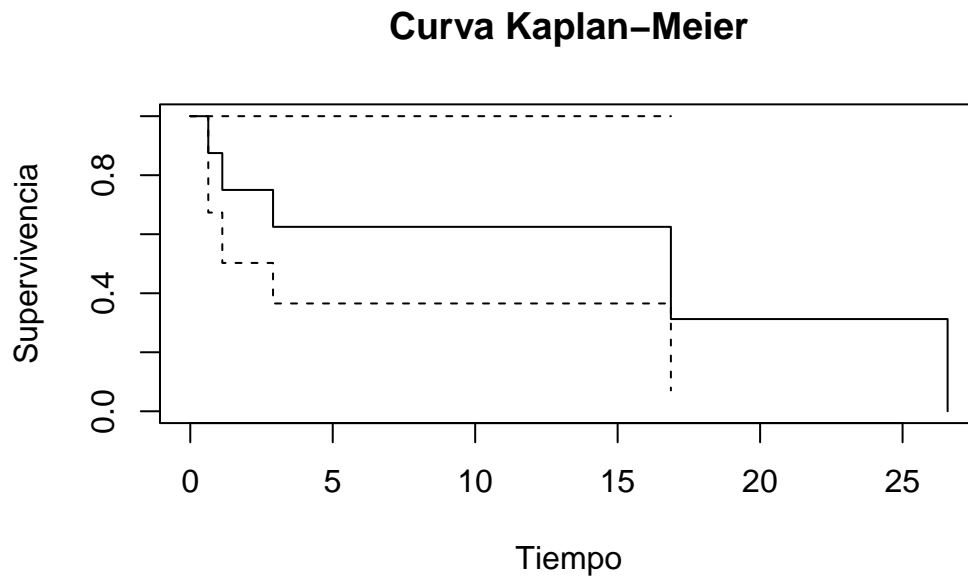
paciente	tiempo	status
1	7	0
2	6	1
3	6	0
4	5	0
5	2	1
6	4	1

- Círculo abierto = censura
- X = evento (muerte)





### 4.3 Visualización de data\_sim



## 5 Referencias

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.