

# Introducción al Análisis de Supervivencia o Tiempo a evento

Jorge Mario Estrada Alvarez PhD. MSc. FETP

# Introducción al análisis de supervivencia

# Analisis de sobrevida ó supervivencia

Conjunto de procedimientos estadísticos aplicados cuando el resultado es el tiempo hasta que ocurre un evento.

Ejemplos de eventos:

- Muerte
- Diagnóstico de enfermedad
- Curación
- Recaída
- Recuperación o alta hospitalaria

# Variable principal

**Variable dependiente:** tiempo hasta el evento (T)

Evento: desenlace de interés, usualmente llamado falla (failure)

Puede representar un desenlace negativo (muerte, enfermedad) o positivo (recuperación).



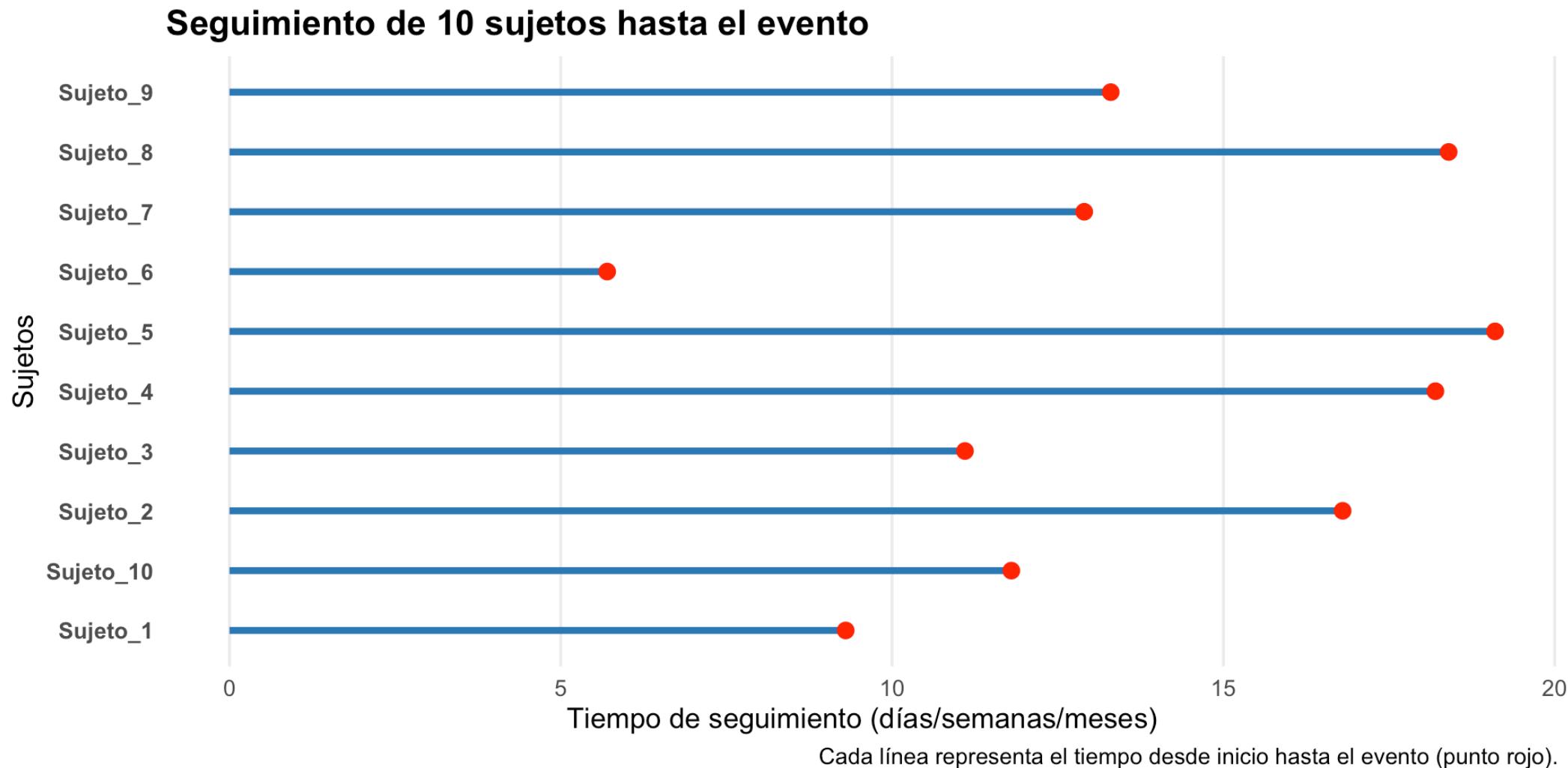
Tip

Idea clave: el tiempo hasta el evento es la información esencial; no solo si ocurre o no.

# Ejemplos de uso de análisis Tiempo a Evento

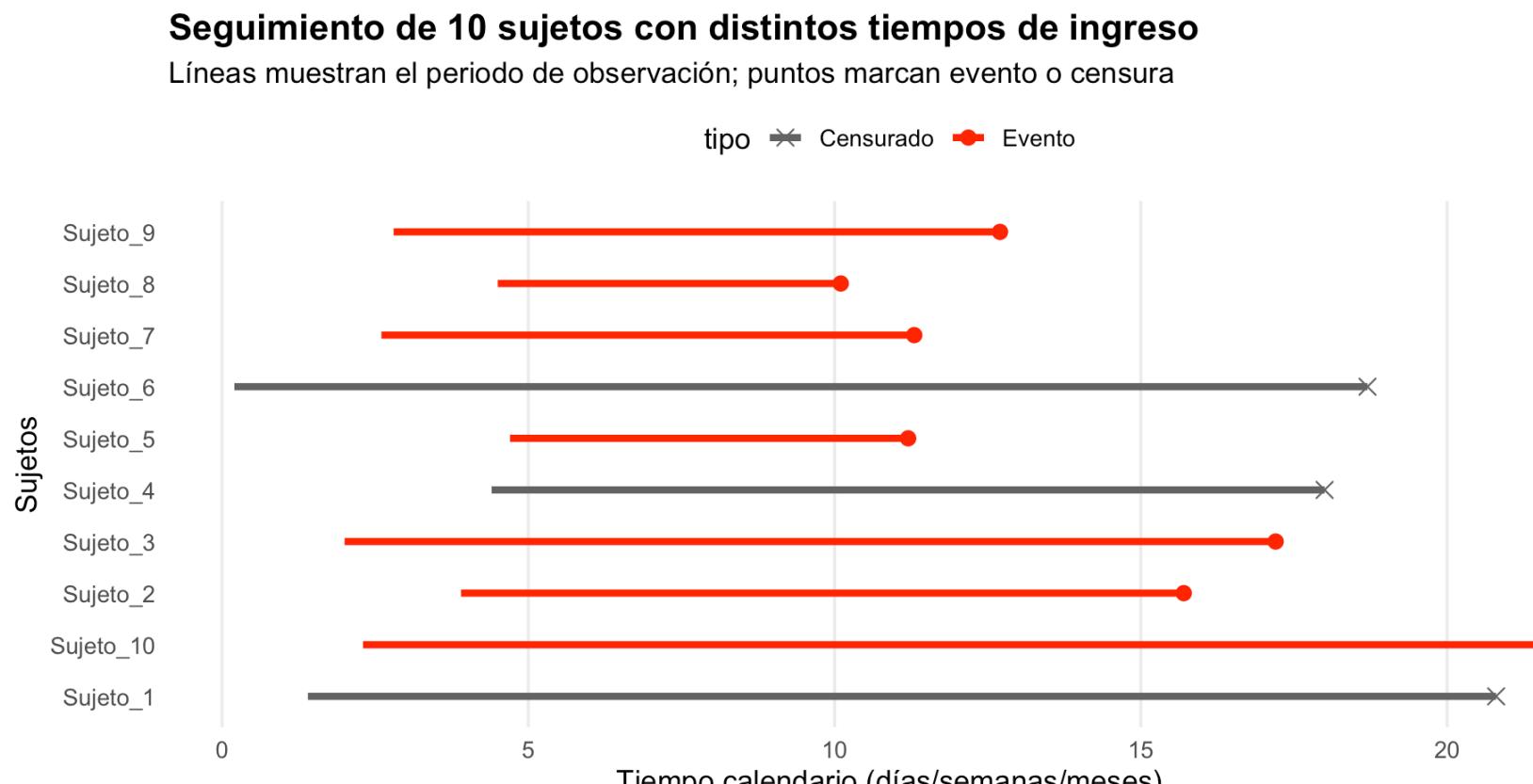
Contexto	Evento	Tiempo
Leucemia	Recaída tras remisión	Semanas
Cohorte libre de ECV	Incidencia de enfermedad cardíaca	Años
Adultos mayores	Muerte	Años
Exreclusos	Reincidencia (rearresto)	Semanas
Trasplante cardíaco	Muerte postoperatoria	Meses

# Datos con seguimientos perfectos y de duración hasta conocer desenlace de todos



# Situación real en seguimientos

- No todos los integrantes se mantienen en seguimiento (perdidas)
- Otros eventos concomitantes afectan su permanencia
- El estudio no puede durar hasta que el ultimo sujeto se conozca su desenlace.
- no todos ingresan en el mismo tiempo calendario.



# ¿Qué es la censura?

La censura ocurre cuando se tiene información parcial sobre el tiempo al evento, pero no el tiempo exacto.

Es un componente esencial del análisis de supervivencia.

Ejemplos:

- El evento no ocurre antes de finalizar el estudio.
- El individuo se pierde en el seguimiento.
- El individuo abandona el estudio o muere (si la muerte no es el evento de interés).

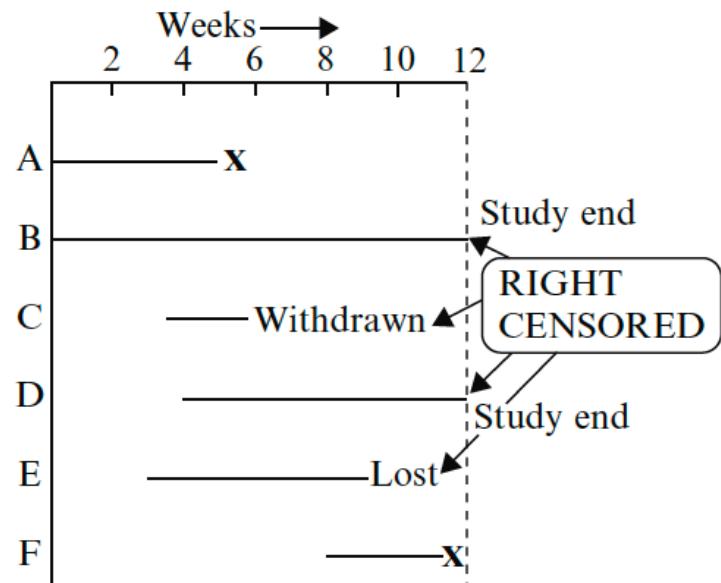
# Tipos de censura

- Censura a la derecha (Right-censored)
- Censura a la izquierda (Left-censored)
- Censura por intervalo (Interval-censored)

# Censura a la derecha (Right-censored)

**Definición:** El evento ocurre después del periodo observado.

**Right-censored:** true survival time is equal to or greater than observed survival time



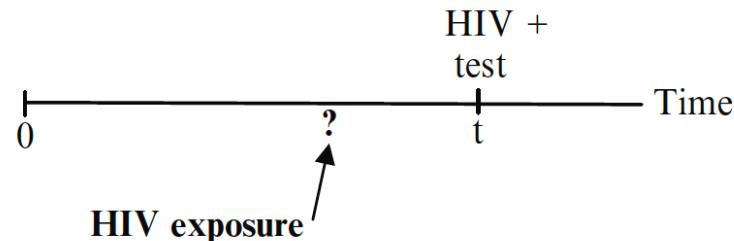
**Ejemplo:**

- Sujeto se retiro del estudio
- Finalizo el seguimiento sin presentar el evento
- Lo afecto otro evento

# Censura a la izquierda (Left-censored)

**Definición:** El evento ocurrió antes del inicio del seguimiento.

**Left-censored:** true survival time is less than or equal to the observed survival time



Kleinbaum & Klein. Survival Analysis – A Self-Learning Text. 3<sup>a</sup> ed.

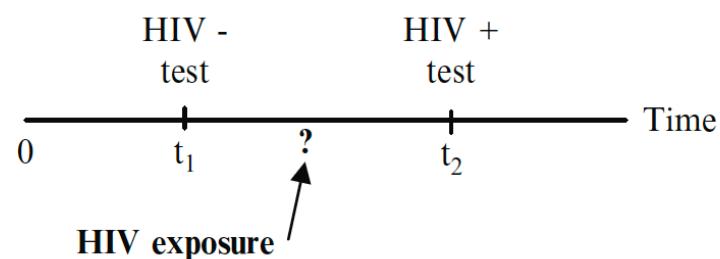
**Ejemplo:**

- Una persona ya era VIH positiva antes del estudio.
- Ingresa al estudio porque se conoce el inicio de síntomas pero no momento de infección

# Censura por intervalo (Interval-censored)

**Definición:** El evento ocurre dentro de un intervalo conocido ( $(t_1, t_2)$ ).

**Interval-censored:** true survival time is within a known time interval



Kleinbaum & Klein. Survival Analysis – A Self-Learning Text. 3<sup>a</sup> ed.

## Ejemplo:

- VIH negativo en  $(t_1)$  y positivo en  $(t_2)$ .
- Se conoce que desarrollo el evento entre dos momentos de seguimiento
- Se conoce dos momentos (fechas) donde mas probable ocurrio la exposición o infección

# Terminología y notación básica

Variable aleatoria de tiempo de supervivencia:

$\backslash(T=\cdot)$  tiempo hasta el evento.

$\backslash(t=\cdot)$  valor específico de  $\backslash(T=\cdot)$

$\backslash(\delta = \cdot)$  indicador del evento (1 = falla, 0 = censura).

# Función de supervivencia

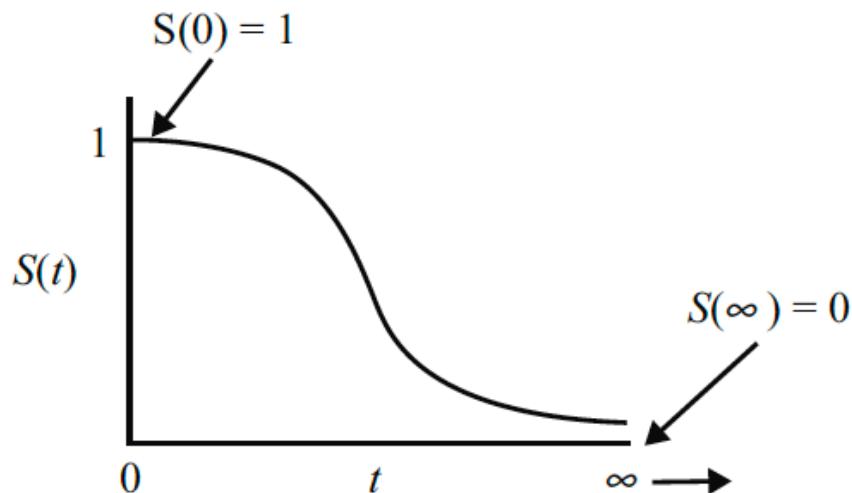
Probabilidad de sobrevivir más allá del tiempo  $t$

$$S(t) = P(T > t)$$

Propiedades:

- $S(0) = 1$
- $S(t)$  disminuye con el tiempo
- Tiende a 0 si todos eventualmente experimentan el evento.

Theoretical  $S(t)$ :



# Función de riesgo o de falla (Hazard)

Mide la tasa instantánea de ocurrencia del evento, dada la supervivencia hasta  $\bar{S}(t)$

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

## Propiedades:

- $h(t) \geq 0$
- No tiene límite superior.
- No es una probabilidad, sino una tasa instantánea condicionada.

También llamada tasa de falla condicional o intensidad de riesgo.

# Relación entre $\backslash(S(t)\backslash)$ y $\backslash(h(t)\backslash)$

$$\begin{aligned} S(t) &= \exp \left[ -\int_0^t h(u) du \right] \\ h(t) &= -\frac{d S(t)}{d t} / S(t) \end{aligned}$$

Si se conoce una de las dos, la otra puede obtenerse.

Función	Enfoca en	Perspectiva
$\backslash(S(t)\backslash)$	Sobrevivir más allá de t	Supervivencia acumulada
$\backslash(h(t)\backslash)$	Riesgo instantáneo de falla	Probabilidad condicional de evento

# Estructura básica de datos para el análisis en software

El formato estándar de los datos para análisis de supervivencia contiene:

Variable	Descripción
<b>id</b>	Identificador del individuo
<b>t</b>	Tiempo observado hasta evento o censura
<b>d</b>	Indicador: 1 = evento, 0 = censura
<b>X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>p</sub></b>	Covariables (tratamiento, edad, sexo, etc.)

# Estimando $S(t)$ : Metodo de Kaplan-Meier

# Formula sin censura y todos al final hacen el evento (condición teorica)

$$[S(t) = P(T > t)] [S(t_i) = \frac{\text{Nº sobrevivientes pasado } t_i}{n}]$$

- Con corrección por censura a derecha:

$$[S(t) = P(T > t) \times P(T \geq t / T > t)] ## Recordar:$$

# Estructura de datos

- Información mínima requerida es Tiempo de observación, datos censurado o no (la estimación mas común es bajo censura a derecha).

Siguientes datos corresponden a un cohorte de individuos con seguimiento a un año:

	<code>id</code>	<code>tiempo</code>	<code>estado</code>
1	1	2	Falla
2	2	3	Falla
3	3	4	Falla
4	4	5	Falla
5	5	6	Falla
6	6	7	Falla
7	7	8	Falla
8	8	9	Falla
9	9	10	Falla
10	10	11	Falla

# Procedimiento para estimación de curva supervivencia K-M

Paso 1: Agrupar los datos por tiempos donde ocurrieron fallos de menor a mayor y cuento cuantas fallas y censuras ocurrieron:

- Datos sin censura

tiempo	n	Falla	s_t
2		1	
3		1	
4		1	
5		1	
6		1	
7		1	
8		1	
9		1	
10		1	
11		1	

Para  $t = 2$  y  $n = 10$

$$S(t>2) = (10-1)/10 = 0.9$$

Para  $t = 3$  y  $n = 10$

$$S(t>3) = (9-1)/10 = 0.8$$

y asi para cada tiempo de falla.

- Datos con censura a derecha

tiempo	n	Falla	Censura	s_t
2		1	0	
3		1	0	
4		0	1	
5		1	0	
6		0	1	
7		1	0	
8		0	1	
9		1	0	
10		0	1	
11		1	0	

Para  $t=2$  como es el primer instante de tiempo y no han ocurrido censuras se inicia como el anterior:

$$S(t>2) = (10-1)/10 = 0.9$$

$$S(t>3) = S(t>2) \times S(t>=2/t>3)$$

$$S(t>3) = 0.9 \times (9-1)/9 = 0.9 \times 0.888 = 0.8$$

...:

$S(t>4) =$  Falla  $(0)$ , Censura  $(1)$ , la sobrevida es la misma del anterior  $(0.8)$

$$S(t>5) = 0.8 \times (7-1)/7 = 0.8 \times 0.857 = 0.6856$$

$S(t>6) =$  Falla  $(0)$ , Censura  $(1)$ , la sobrevida es la misma del anterior  $(0.6856)$

$$S(t>7) = 0.6856 \times (5-1)/5 = 0.6856 \times 0.8333 = 0.5484$$

$S(t>8) =$  Falla  $(0)$ , Censura  $(1)$ , la sobrevida es la misma del anterior  $(0.5484)$

$$S(t>9) = 0.5484 \times (3-1)/3 = 0.5484 \times 0.6666 = 0.3655$$

$S(t>10) =$  Falla  $(0)$ , Censura  $(1)$ , la sobrevida es la misma del anterior  $(0.3655)$

$$S(t>11) = 0.3655 \times (1-1)/1 = 0$$

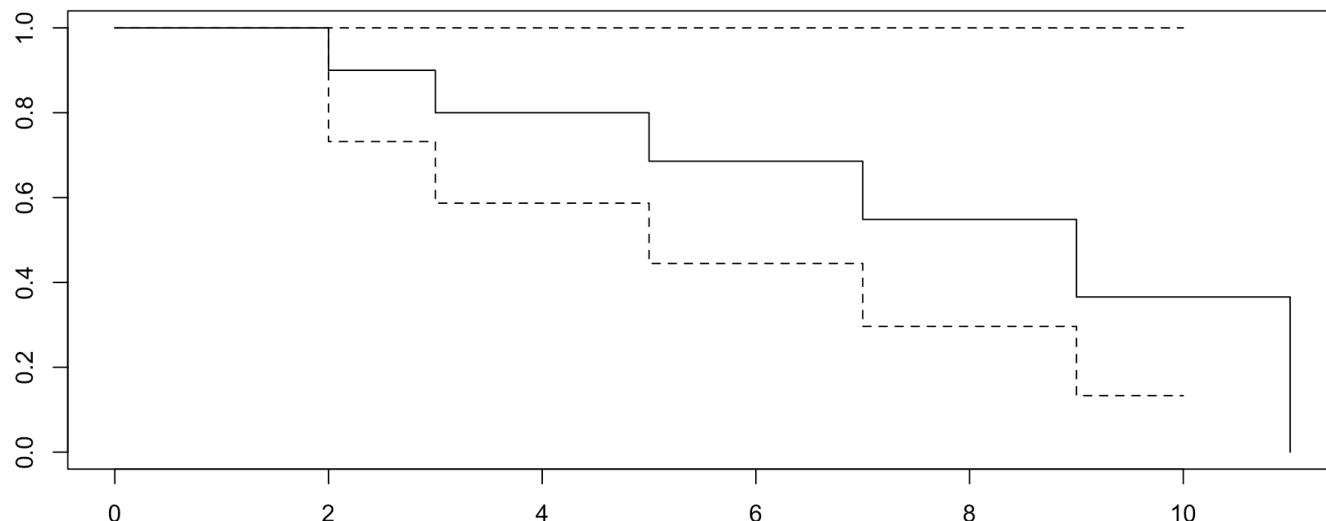
Finalmente la tabla de sobrevida quedaría así:

j	\	Censuras	\	\	Factor \	\
	(t_{(j)})\)	(n_j\)	(d_j\)	((n_j - d_j)/n_j\)		(S(t_{(j)}))\)
1	2	0	10	1	\frac{9}{10}	0.9
2	3	0	9	1	\frac{8}{9} \times 0.9 = 0.888	0.8
3	5	1	7	1	\frac{6}{7} \times 0.8 = 0.6857	0.6857

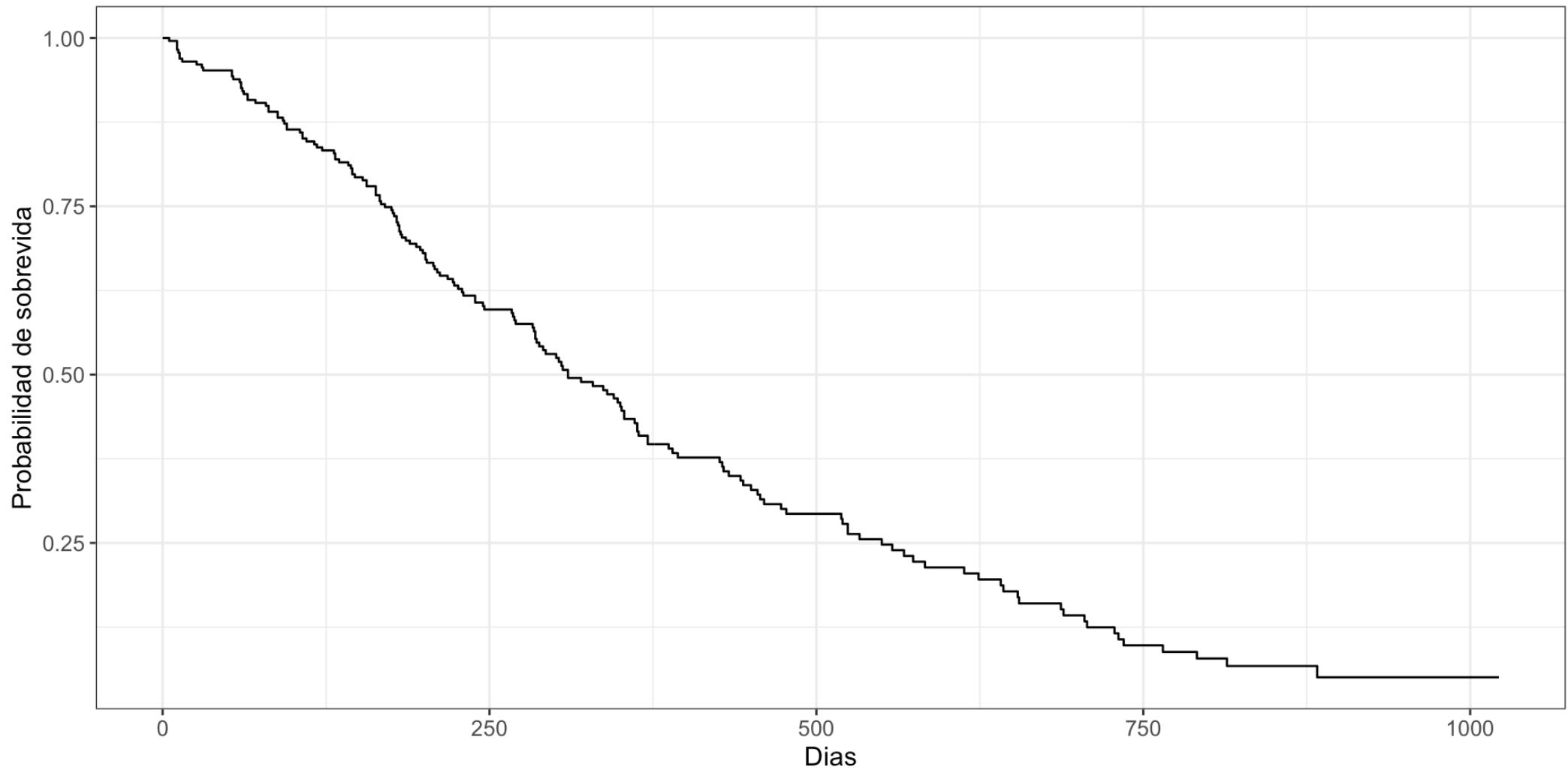
j	\	Censuras	\	\	Factor \	\
	(t_{(j)})\)		(n_j\)	(d_j\)	((n_j- d_j)/n_j\)	(S(t_{(j)}))\)
4	7	1	5	1	\ \\ \frac{4}{5} \times 0.8 \)	0.6857 \\ 0.5486
5	9	1	3	1	\ \\ \frac{2}{3} \times 0.6666 \)	0.3657 \\ 0.5486 \\ 0.3657
6	11	1	1	1	\ \\ \frac{0}{1} \times 0 \)	0.3657 \\ 0 \\ 0

```
Call: survfit(formula = Surv(tiempo, estado) ~ 1, data = df)
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
2	10	1	0.900	0.0949		0.732		1
3	9	1	0.800	0.1265		0.587		1
5	7	1	0.686	0.1515		0.445		1
7	5	1	0.549	0.1724		0.296		1
9	3	1	0.366	0.1884		0.133		1
11	1	1	0.000	NaN		NA		NA



# Interpretaciones de la Curva K-M



# Estimaciones de supervivencia para tiempos específicos

```
1 summary(survfit(Surv(time, status) ~ 1, data = lung), times = c(
```

```
Call: survfit(formula = Surv(time, status) ~ 1, data = lung)
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
182	156	66	0.708	0.0303		0.651		0.770
365	65	55	0.409	0.0358		0.345		0.486

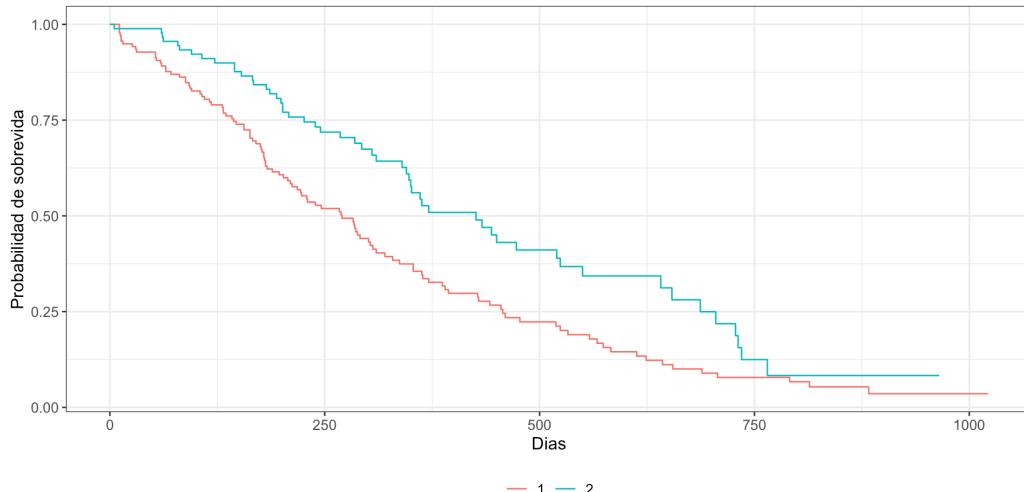
```
1 survfit(Surv(time, status) ~ 1, data = lung)
```

```
Call: survfit(formula = Surv(time, status) ~ 1, data = lung)
```

	n	events	median	0.95LCL	0.95UCL
[1, ]	228	165	310	285	363

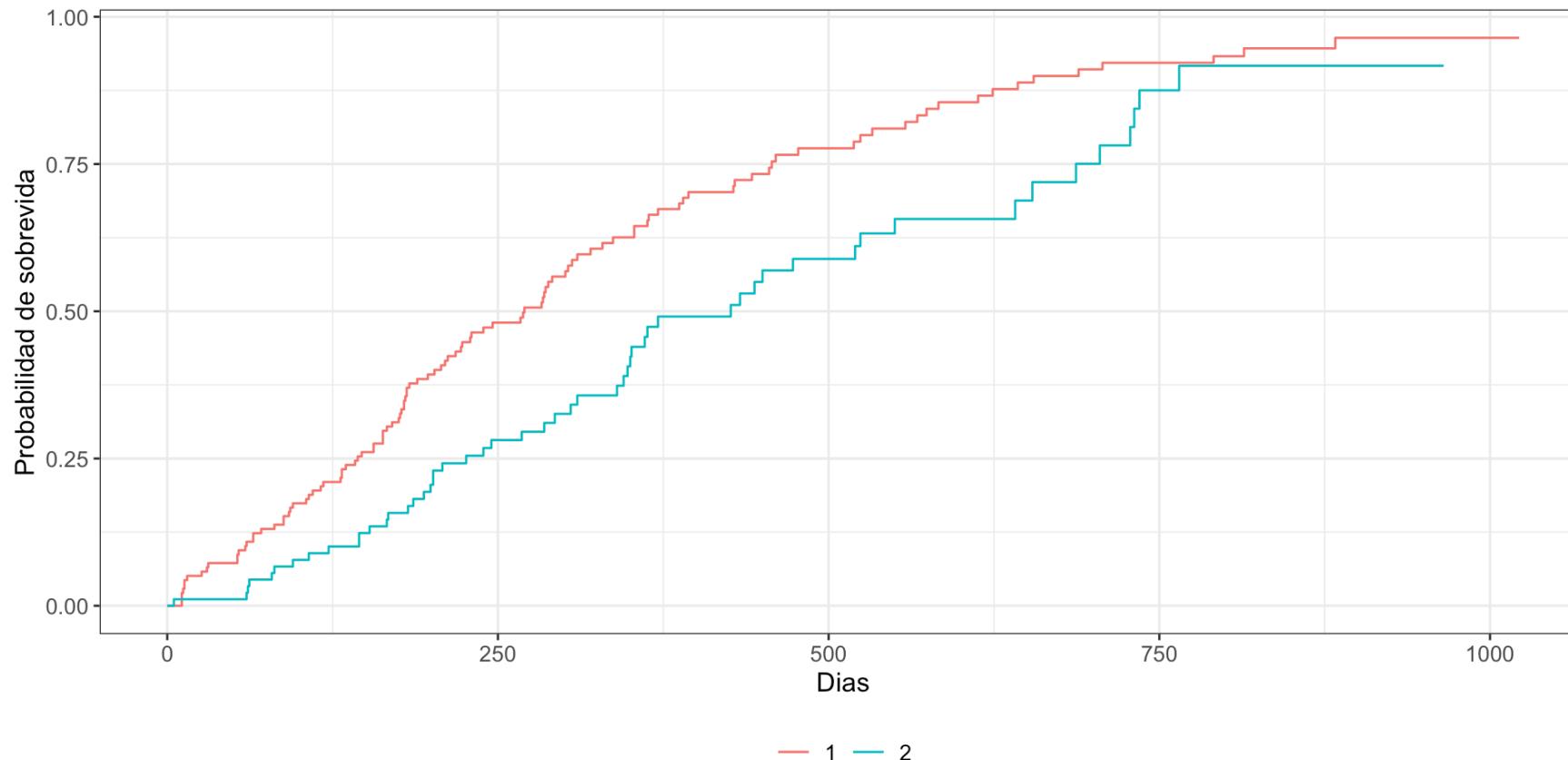
# Comparando Curvas de supervivencia: Log-Rank Test

```
Call:  
survdiff(formula = Surv(time, status) ~ sex, data = lung)  
  
          N Observed Expected (O-E)^2/E (O-E)^2/V  
sex=1 138      112    91.6     4.55     10.3  
sex=2  90       53    73.4     5.68     10.3  
  
Chisq= 10.3 on 1 degrees of freedom, p= 0.001
```



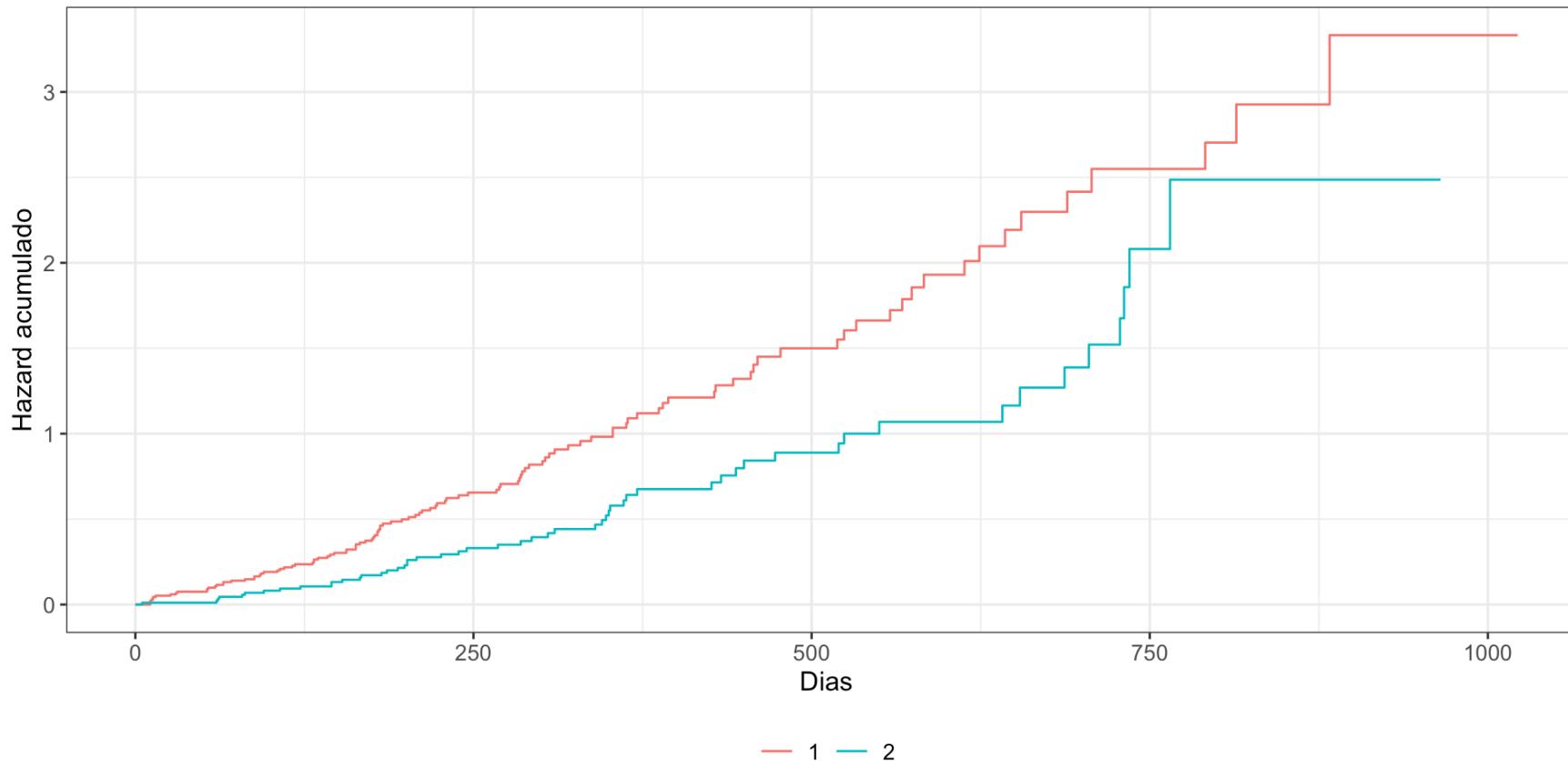
# Otras funciones: Proporción acumulada del evento

- corresponde al complemento de  $S(t)$  es decir  $(1-S(t))$



# Otras funciones: Hazard Acumulado o riesgo Acumulado

- corresponde  $\lambda(H(t) = -\log(S(t))$



# Resumen

Función	Qué mide	Unidad conceptual	Intuición práctica
$\bar{S}(t)$	Probabilidad de seguir libre del evento	Probabilidad $(0-1)$	“Supervivencia”
$\bar{(1 - S(t))}$	Proporción que ya tuvo el evento	Probabilidad $(0-1)$	Incidencia acumulada
$\bar{h}(t)$	Tasa instantánea de evento	$(1/\text{tiempo})$	Velocidad del riesgo
$\bar{H}(t)$	Riesgo acumulado total	Riesgo adimensional	“Cantidad de riesgo acumulada”