



Pontificia Universidad  
**JAVERIANA**  
Bogotá

MAE STRÍA EN   
**EPIDEMIOLOGÍA**  
CLÍNICA

**BIOESTADÍSTICA AVANZADA**

## **MÓDULO III**

**Semana 11**  
Análisis de Supervivencia

# Material de contenido y aplicación

Martín Alonso Rondón Sepúlveda

## Introducción

El **análisis de supervivencia** es un conjunto de técnicas estadísticas que permiten analizar el tiempo que transcurre hasta que ocurre un evento de interés (por ejemplo, muerte, recaída, falla de un dispositivo, alta hospitalaria, etc.).

Una característica que tiene este análisis es que no siempre, todos los individuos observados presentan el evento de interés durante el período de seguimiento, concepto conocido como dato censurado. La no presencia del evento de interés se puede dar por diferentes circunstancias, como por ejemplo, que el individuo nunca presentó o desarrolló el evento de interés, o por pérdida de seguimiento o porque falleció por causas diferentes a las de interés de investigación (por ejemplo, en un accidente de tránsito).

## Objetivos de análisis de supervivencia

1. Estimar la probabilidad de supervivencia: Calcular la probabilidad de que un individuo sobreviva hasta un tiempo determinado sin experimentar el evento de interés.
2. Comparar supervivencia entre grupos: Comparar las funciones de supervivencia de diferentes grupos (por ejemplo, pacientes con tratamiento A vs. tratamiento B) para ver si hay diferencias significativas.
3. Identificar factores de riesgo: Determinar qué variables (edad, sexo, tratamiento, etc.) afectan el tiempo hasta que ocurra el evento, utilizando modelos como el de regresión de Cox.

4. Estimar la duración media de supervivencia: Calcular la mediana o la media del tiempo hasta el evento, lo que ayuda a entender el comportamiento general de la población.
5. Predecir la supervivencia futura: Usar modelos para prever la probabilidad de supervivencia de un individuo en el futuro, basado en sus características.
6. Evaluar la efectividad de un tratamiento: Determinar si un tratamiento o intervención mejora la supervivencia en comparación con otros tratamientos o un grupo de control.

## Conceptos básicos de análisis de supervivencia

### Evento de interés

Suceso cuya ocurrencia se estudia en el análisis de supervivencia (por ejemplo, muerte, recaída, etc.). El evento de interés suele denominarse falla o fracaso, aunque no siempre implica un resultado negativo. Por ejemplo, puede ser la recaída de una enfermedad o el regreso al trabajo después de una cirugía.

### Variable desenlace: Tiempo de supervivencia

La variable desenlace (notada como  $T$ ), es una variable aleatoria continua y no negativa que representa el tiempo desde el inicio del seguimiento (por ejemplo diagnóstico de la enfermedad, ingreso al estudio o inicio del tratamiento) hasta la ocurrencia del evento de interés.

- $t$ : valor específico observado o a evaluar de  $T$ .

Representa un tiempo particular en el que se quiere evaluar la probabilidad de supervivencia o de ocurrencia de un evento. Por ejemplo, si se quiere evaluar si un proceso de tratamiento con quimioterapia ayuda a un paciente a sobrevivir 5 años ó más desde el momento del inicio del tratamiento, la siguiente sería la notación:  $P(T \geq 5)$

## Censura y truncamiento

En el análisis de supervivencia, la censura y el truncamiento son mecanismos mediante los cuales la información sobre el tiempo al evento es incompleta. Aunque ambos implican pérdida parcial de información, difieren en el momento en que ocurre esta limitación respecto al período de observación del individuo.

La censura ocurre cuando un individuo forma parte del estudio, pero el tiempo exacto en el que ocurre el evento de interés no se observa completamente. En estos casos, el individuo contribuye con información parcial al análisis, ya que se conoce que el evento ocurrió antes o después de cierto momento, pero no el tiempo exacto. Específicamente tenemos:

### Censura a la derecha

La *censura a la derecha* ocurre cuando el evento de interés no se observa durante el período de seguimiento del individuo. En estas situaciones, solo se sabe que el tiempo real hasta el evento es mayor que el último tiempo en el que el individuo fue observado.

Esto puede suceder, por ejemplo, cuando:

- El estudio finaliza antes de que ocurra el evento de interés.
- El individuo se pierde durante el seguimiento y no se dispone de información posterior.
- El participante se retira voluntariamente del estudio antes de que ocurra el evento.
- El individuo presenta un evento que impide continuar su seguimiento para el evento de interés (por ejemplo, fallece por una causa diferente).

### Censura a la izquierda

Ocurre cuando el evento de interés ya ha sucedido antes del inicio del período de observación, pero no se conoce el momento exacto en que ocurrió.

### Censura por intervalo

Ocurre cuando se sabe que el evento ocurrió dentro de un intervalo de tiempo, pero no se conoce el momento exacto dentro de ese intervalo. Esto sucede, por ejemplo, cuando los individuos son evaluados en visitas periódicas y el evento se detecta entre dos evaluaciones consecutivas.

El truncamiento por su parte sucede cuando ciertos individuos no son incluidos en la muestra porque su tiempo al evento queda fuera del intervalo de observación definido por el estudio. Es decir, el investigador solo observa individuos cuyos tiempos cumplen una condición determinada, por lo que algunos casos potenciales nunca llegan a formar parte de los datos analizados.

Específicamente también tenemos:

## Truncamiento a la izquierda

Ocurre cuando solo se incluyen en el estudio los individuos cuyo tiempo al evento es mayor que un cierto punto de entrada. Es decir, los individuos cuyo evento ocurrió antes de ese momento no son observados ni registrados en la muestra.

- **Ejemplo:** en un estudio sobre la supervivencia de pacientes tras un trasplante de hígado, solo se incluyen en la cohorte aquellos que llegaron vivos al hospital de referencia donde comienza el seguimiento. Los pacientes que fallecieron antes de llegar al hospital (aunque hubieran sido candidatos al trasplante) no son considerados en el análisis.

## Truncamiento a la derecha

Ocurre cuando solo se incluyen individuos cuyo tiempo al evento es menor que un cierto límite superior. Los individuos cuyo evento ocurre después de ese punto no son incluidos en el estudio.

- **Ejemplo:** en un estudio sobre la supervivencia de pacientes tras un trasplante de hígado, solo se incluyen aquellos que fallecen dentro de los primeros 5 años posteriores al trasplante. Los pacientes que sobreviven más allá de esos 5 años no son considerados en el análisis.

## Truncamiento por intervalo

Solo se observan individuos cuyo evento ocurre dentro de un rango específico de tiempos.

- **Ejemplo:** en un estudio sobre la aparición de complicaciones cardiovasculares después de iniciar tratamiento para hipertensión, solo se incluyen pacientes cuyo evento (infarto o accidente cerebrovascular) ocurre entre 2 y 5 años de seguimiento, eso significa que se excluyen los que presentan el evento antes de los 2 años y los que lo presentan después de 5 años.

## Ejemplo de los conceptos básicos

En un estudio sobre la supervivencia de pacientes con cáncer, el evento de interés es la muerte por cáncer. La variable desenlace es el tiempo desde el diagnóstico hasta la muerte.

En este estudio se utiliza un registro hospitalario en el que los pacientes son incluidos cuando acuden por primera vez al hospital después de haber sido diagnosticados. Debido a esto, algunos pacientes pueden haber sido diagnosticados antes de entrar al estudio. Esta situación se conoce como truncamiento a la izquierda, ya que solo se observan pacientes que han sobrevivido desde el diagnóstico hasta el momento en que ingresan al registro.

Una vez que los pacientes entran al estudio, son seguidos durante un periodo de tiempo para observar si ocurre el evento de interés. Sin embargo, al finalizar el estudio algunos pacientes pueden continuar vivos o perderse durante el seguimiento. En estos casos, el tiempo exacto hasta la muerte no se observa y se dice que el dato está censurado a la derecha, ya que solo se sabe que el tiempo de supervivencia es mayor que el tiempo observado.

Por lo tanto, cada paciente incluido en el estudio está asociado a un tiempo observado y a un indicador de evento ( $\delta$ ) así:

$$\delta = \begin{cases} 1 & \text{si el evento de interés ocurre} \\ 0 & \text{si el dato está censurado} \end{cases}$$

En este caso el evento de interés es el fallecimiento del paciente por cáncer, entonces:

$\delta=1$  → El paciente falleció de cáncer durante el tiempo de estudio.

$\delta=0$  → El paciente no falleció por cáncer durante el período observado (censura a la derecha).

Específicamente, en este estudio los datos presentan truncamiento a la izquierda, porque solo se observan pacientes que sobrevivieron hasta el momento de su inclusión en el registro, y censura a la derecha, porque algunos pacientes continúan vivos al finalizar el seguimiento.

## Función de supervivencia

La función de supervivencia, denotada como

$$S(t)$$

, es una función que describe la probabilidad de que un individuo sobreviva más allá de un tiempo específico

$$t$$

. Matemáticamente, se define como:

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

donde  $F(t)$  es la función de distribución acumulada de  $T$ , que representa la probabilidad de que el evento ocurra en o antes del tiempo  $t$ .

$t$	$S(t)$
1	$S(1) = P(T > 1)$
2	$S(2) = P(T > 2)$
3	$S(3) = P(T > 3)$
$\vdots$	$\vdots$

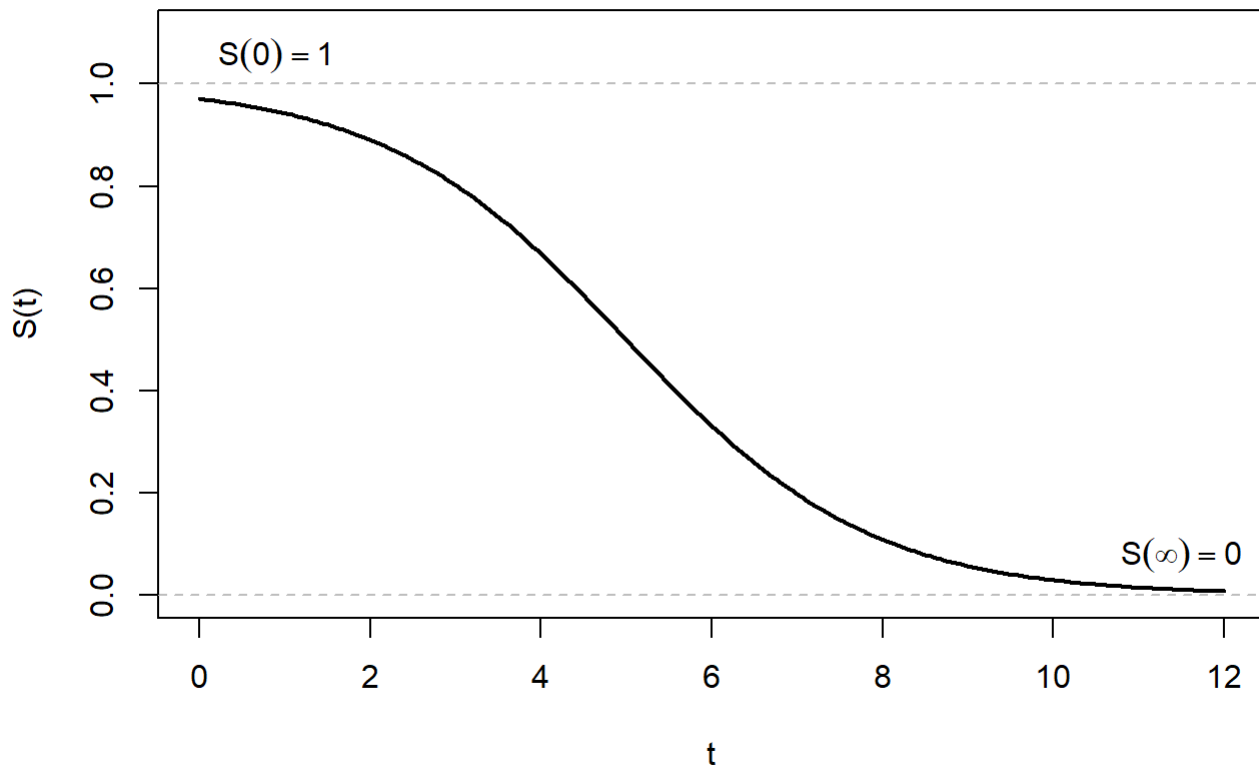
## Función de supervivencia (curva teórica)

Desde el punto de vista teórico, la función de supervivencia es una curva  $[0, \infty)$ , que inicia con una probabilidad de 1 (en el tiempo 0) y va disminuyendo a medida que el tiempo avanza, acercándose a 0 a medida que

$$t$$

tiende a infinito.

## Función de Supervivencia



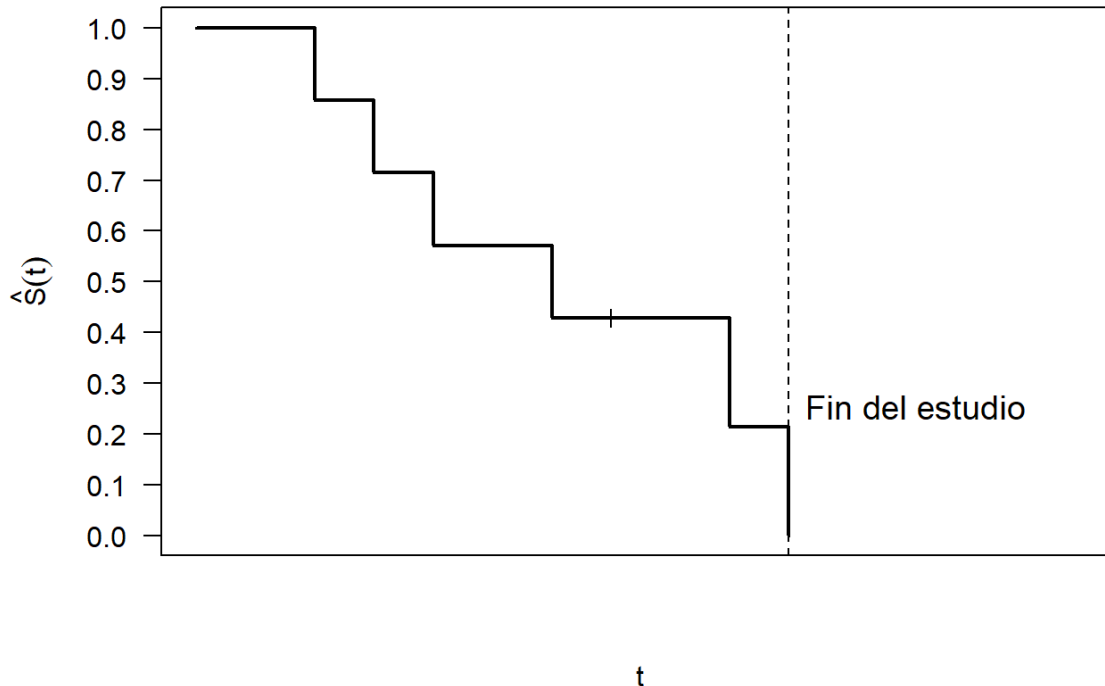
### Características:

- En el momento  $t = 0, S(t) = S(0) = 1$
- Es decreciente (inicia en 1 y van en descenso)
- En momento:  $t = \infty, S(t) = S(\infty) = 0$

Por ejemplo, si el evento es muerte, y el período de estudio es muy extenso, al final todos van a fallecer.

## Función de supervivencia (curva en la práctica)

Función de Supervivencia



- Casi siempre es una 'escalera'
- Curva supervivencia estimada  $\hat{S}^{\wedge}$
- El tiempo seguimiento no es infinito y pueden existir riesgos competitivos
- Es común que  $\hat{S}^{\wedge}$  no tome el valor de 0 al final del estudio

*En resumen tenemos:*

- $T$  = tiempo de supervivencia de variable aleatoria
- $t$  = valor específico de  $T$
- $d$  = variable para indicar falla (evento) o censura (1,0)
- $S(t)$  = función de supervivencia

## Funciones $S(t)$ cuando no hay censuras

*Datos que se necesitan:*

- Tiempo de seguimiento: Generalmente se calcula considerando la siguiente información:
  - Fecha diagnóstico
  - Fecha de “falla”
  - Fecha de último contacto de los que se perdieron de seguimiento o están sin evento al final
- Causa del evento

## Ejemplo de aplicación cuando no hay censuras

Un estudio quería evaluar cuánto tiempo sobreviven los pacientes después de haber sido diagnosticados con cáncer y si existen diferencias en la supervivencia entre hombres y mujeres. Para ello, tomaron una muestra representativa de pacientes con el propósito de aproximarse a la distribución de los tiempos de supervivencia. Los siguientes fueron los resultados obtenidos

**Información de los tiempos hasta el evento de los pacientes con cáncer**

id	t	d	sexo
1	37	1	F
2	680	1	M
3	250	1	F
4	250	1	F
5	377	1	F
6	840	1	F
7	847	1	M
8	532	1	M
9	670	1	F

id	t	d	sexo
10	468	1	F
11	432	1	F
12	215	1	F

Donde:

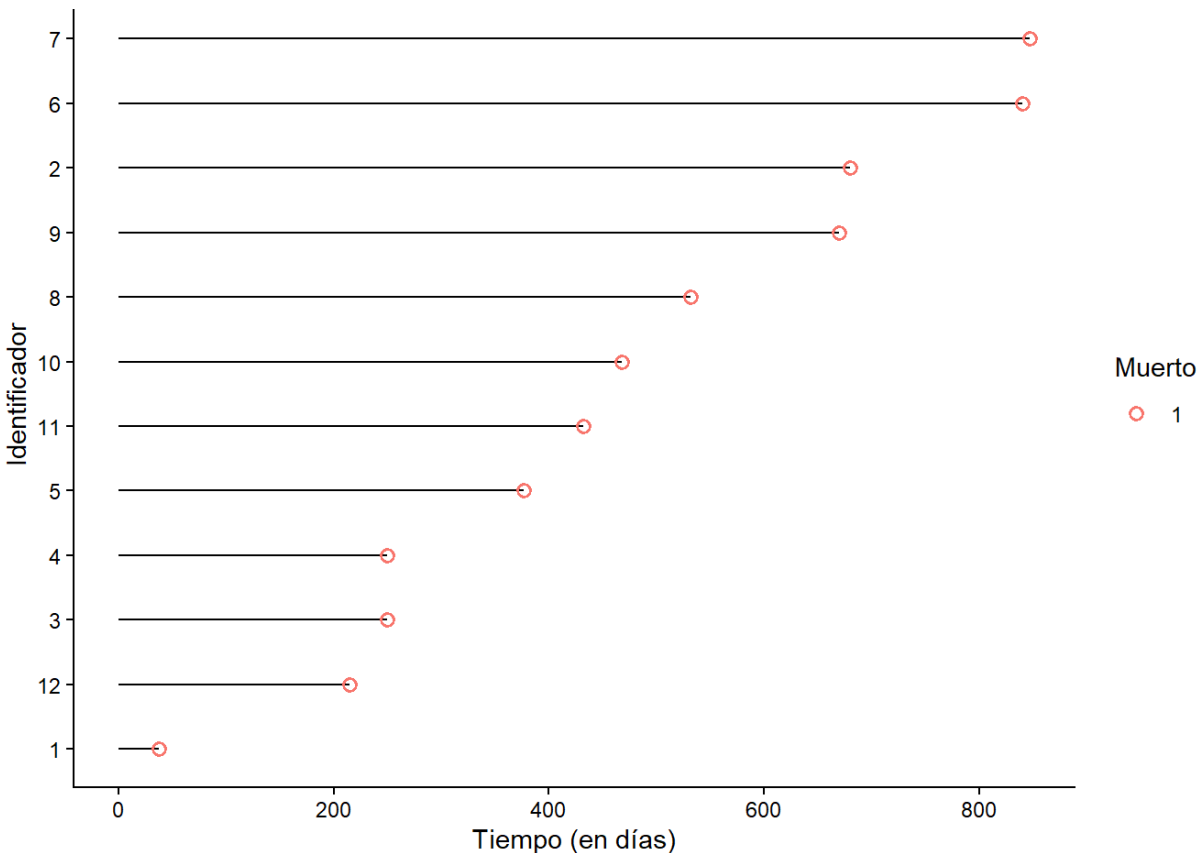
*id*: Identificador del paciente.

*t*: Tiempo de supervivencia en días desde el inicio del seguimiento hasta la ocurrencia del evento.

*d*: Indicador de evento (1 = el evento ocurrió, 0 = censurado).

*sexo*: Sexo del paciente (F. Femenino M. Masculino).

## Visualización de la información:



El gráfico presentado muestra los tiempos de supervivencia individuales de los pacientes incluidos en el estudio, donde cada línea horizontal representa el periodo de seguimiento de un paciente desde el inicio hasta la ocurrencia del evento de interés, marcado con un círculo rojo.

En este gráfico podemos observar que todos los pacientes fallecieron durante el periodo de observación, por lo que no existen datos censurados. Se observa una amplia variabilidad en los tiempos de supervivencia, que van desde muertes tempranas, como la del paciente 1 en torno a los 37 días, hasta casos de mayor duración, como los pacientes 11 y 12 que superan los 800 días. A partir de esta información queremos calcular la función de supervivencia.

## Estrategias para estimar la función de supervivencia:

- **Modelos paramétricos:** suponen que los tiempos de supervivencia siguen una distribución conocida (exponencial, Weibull, log-normal, etc.).
- **Modelos semiparamétricos:** como el modelo de Cox, que permite incorporar covariables sin especificar la forma de la función de riesgo basal.
- **Métodos no paramétricos:** como el estimador de **Kaplan–Meier**, que no requiere asumir una distribución previa y es el más utilizado para describir los datos de forma exploratoria. Este método estima la probabilidad de supervivencia acumulada y permite visualizar qué grupo tiene mejor pronóstico.

A continuación, revisaremos como estimar la función de supervivencia a través del método de Kaplan–Meier.

### Fórmula del estimador de Kaplan–Meier

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

Donde:

$t_j$ : tiempos de evento,

$n_j$ : número en riesgo en  $t_j$ ,

$d_j$ : número de eventos en  $t_j$ .

## Método de Kaplan–Meier para datos sin censuras

### 1. Ordenar los tiempos de evento:

Colocar todos los tiempos de supervivencia en orden cronológico.

id	t	d	sexo
1	37	1	F
12	215	1	F
3	250	1	F
4	250	1	F
5	377	1	F
11	432	1	F
10	468	1	F
8	532	1	M
9	670	1	F
2	680	1	M
6	840	1	F
7	847	1	M

## 2. Hacer la tabla con las estimaciones

Para crear la tabla necesitamos:

$t_i$	$N_i$ (riesgo)	$d_i$ (fallas)	$S_i$ (sobrev.)	$S_i / N_i$	Supervivencia
$t_1$					
$t_2$					
...					
$t_j$					

- $N_i$ : Número de pacientes *a riesgo* en cada tiempo de evento  $t_i$  y que corresponde a los sujetos que aún no han experimentado el evento ni están censurados justo antes de  $t_i$ .
- $d_i$ : pacientes con *falla* en cada tiempo  $t_i$ .
- $S_i$ : pacientes que han sobrevivido después del tiempo  $t_i$ : Recuerde que  $d_j = N_i - S_i$
- Calcular la probabilidad condicional de supervivencia en cada  $t_j$ :

$$p_j = 1 - \frac{d_j}{n_j}$$

- Calculando la supervivencia acumulada:

$$\hat{S}(t) = \prod_{t_j \leq t} p_j$$

## 3. Construyendo la tabla (sin repetir valores)

Inicialmente colocamos el tiempo 0 (tiempo en que comienza el estudio, y en el cual todos están vivos)

Posteriormente, colocamos en la tabla el tiempo en el cual se presentó la primera falla (es decir, la primera muerte, que en este ejemplo corresponde a un tiempo de 37 días).

- $N_1$ : 12
- $d_1$ : 1
- $s_1$ : 11
- **Probabilidad condicional de supervivencia en  $t_1$ :**

$$p_1 = 1 - \frac{d_1}{n_1} = 1 - \frac{1}{12} = 0.9167$$

- **Supervivencia acumulada:**

$$\hat{S}(1) = \prod_{t_1 \leq t} p_1 = p_0 * p_1 = 1.000 * 0.9167$$

A continuación, colocamos en la tabla el siguiente menor tiempo en el cual se presentó la falla (es decir, la segunda muerte, que en este ejemplo corresponde a un tiempo de 215 días).

- $N_2$ : 11
- $d_2$ : 1
- $s_2$ : 10
- **Probabilidad condicional de supervivencia en  $t_2$ :**

$$p_2 = 1 - \frac{d_2}{n_2} = 1 - \frac{1}{11} = 0.9091$$

- **Supervivencia acumulada:**

$$\hat{S}(2) = \prod_{t_2 \leq t} p_2 = p_1 * p_2 = 0.9167 * 0.9091 = 0.8334$$

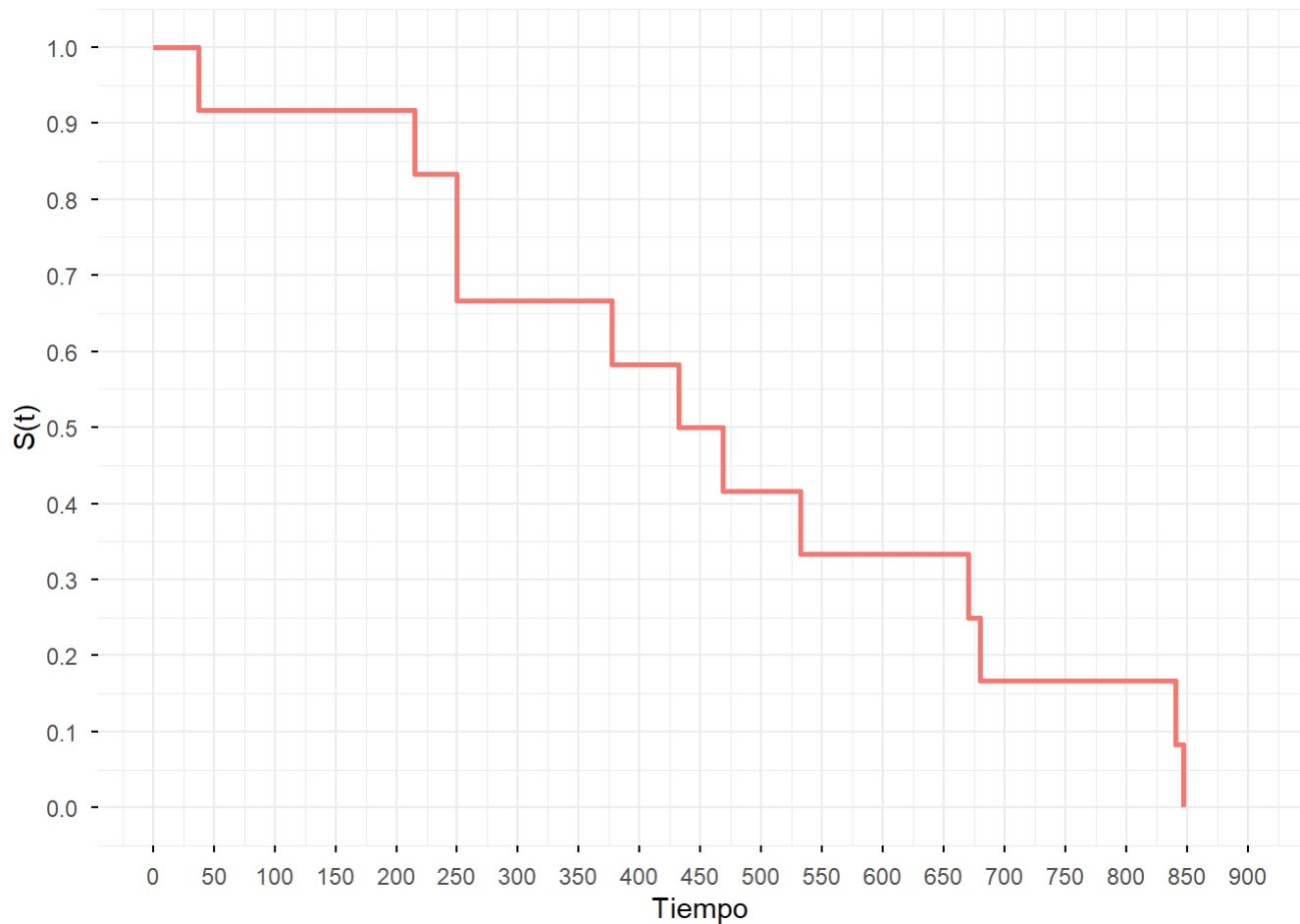
y así sucesivamente hasta obtener la siguiente tabla:

$t_i$	$N_i$ (riesgo)	$\delta_i$ (fallas)	$S_i$ (sobrev.)	$S_i / N_i$	Supervivencia
$t_0=0$	12	0	12	$12/12=1.000$	1.000
$t_1=37$	12	1	11	$11/12=0.917$	0.917
$t_2=215$	11	1	10	$10/11=0.917$	0.834
$t_3=250$	10	2	8	$8/10=0.800$	0.667
$t_4=377$	8	1	7	$7/8 =0.875$	0.584
$t_5=432$	7	1	6	$6/7 =0.857$	0.500
$t_6=468$	6	1	5	$5/6 =0.833$	0.417
$t_7=532$	5	1	4	$4/5 =0.800$	0.333
$t_8=670$	4	1	3	$3/4 =0.750$	0.250
$t_9=680$	3	1	2	$2/3 =0.667$	0.167
$t_{10}=840$	2	1	1	$1/2 =0.500$	0.083
$t_{11}=847$	1	1	0	$0/1 =0.000$	0.000

*Nota:* En el tiempo 250 hubo dos eventos, pero solo se incluye el tiempo una sola vez en la tabla y se especifica que hubo dos eventos

## 4. Graficar la curva de supervivencia

A partir de la tabla anterior obtenemos el siguiente gráfico:



Podemos observar que la curva de supervivencia desciende en cada tiempo donde ocurre un evento (muerte), reflejando la disminución en la probabilidad de supervivencia a lo largo del tiempo.

## Realizando la curva de Kaplan Meier en R

### Paso 1. Crear la base de datos

Ejecute el siguiente código:

```
# Primero creamos los vectores
id <- c(1,2,3,4,5,6,7,8,9,10,11,12)
t <- c(37,215,250,250,377,432,468,532,670,680,840,847)
d <- c(1,1,1,1,1,1,1,1,1,1,1,1)
sexo <- c("F","F","F","F","F","F","F","M","F","M","F","M")

# Posteriormente, creamos la base de datos
datos <- data.frame(ID=id, Tiempo=t, Evento=d, Sexo=sexo)
head(datos)
```

##	ID	Tiempo	Evento	Sexo
## 1	1	37	1	F
## 2	2	215	1	F
## 3	3	250	1	F
## 4	4	250	1	F
## 5	5	377	1	F
## 6	6	432	1	F

### Paso 2. Crear la tabla de Kaplan-Meier

```
km <- survfit(Surv(Tiempo, Evento) ~ 1, data = datos)
km
```

```
## Call: survfit(formula = Surv(Tiempo, Evento) ~ 1, data = datos)
##
##           n events median 0.95LCL 0.95UCL
## [1,] 12      12      450      250      NA
```

### Paso 3. Visualizar tiempos y función de supervivencia en la tabla de Kaplan-Meier

```

tabla <- data.frame(
  Tiempo = km$time,
  En_Riesgo = km$n.risk,
  Fallas = km$n.event,
  Supervivencia = round(km$surv, 3)
)

knitr::kable(tabla, caption = "Tabla de Kaplan-Meier sin censuras")
  
```

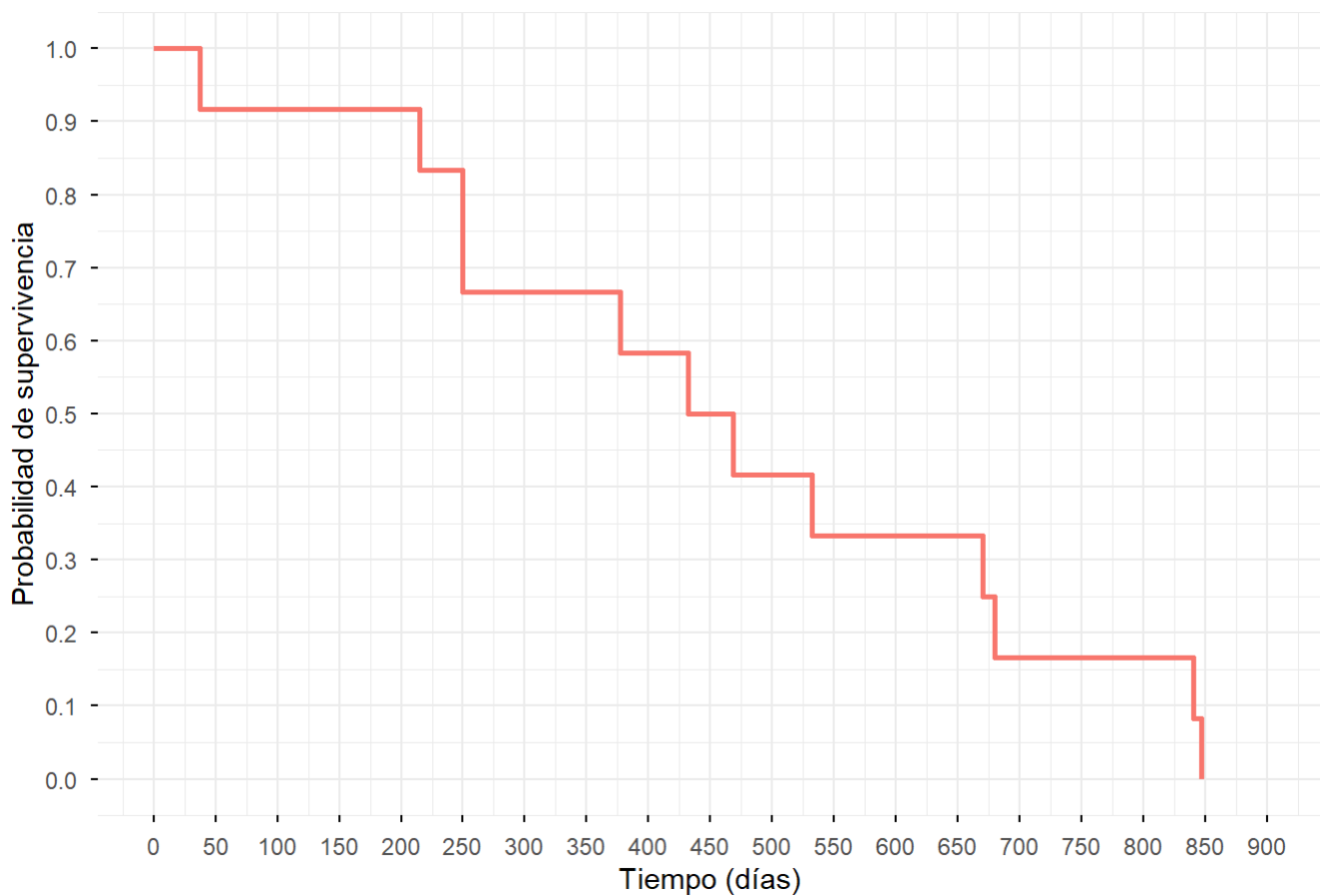
Tabla de Kaplan-Meier sin censuras

Tiempo	En_Riesgo	Fallas	Supervivencia
37	12	1	0.917
215	11	1	0.833
250	10	2	0.667
377	8	1	0.583
432	7	1	0.500
468	6	1	0.417
532	5	1	0.333
670	4	1	0.250
680	3	1	0.167
840	2	1	0.083
847	1	1	0.000

## Paso 4. Creando la curva de Supervivencia

```
ggsurvplot(km, data=datos,  
           conf.int = FALSE,  
           break.time.by = 50,  
           break.y.by = 0.1,  
           xlim = c(0, 900),  
           ggtheme = theme_minimal(),  
           xlab = "Tiempo (días)",  
           ylab = "Probabilidad de supervivencia",  
           title = "Gráfico de Kaplan-Meier sin censuras",  
           legend = "none")
```

Gráfico de Kaplan-Meier sin censuras



## Ejemplo de aplicación del Método de Kaplan–Meier para datos con censuras

Tenemos el segundo conjunto de datos

```
id <- c(1,12,4,3,5,11,10,8,9,2,6,7)
t <- c(37,215,250,250,377,432,468,532,670,680,840,847)
d <- c(1,0,0,1,1,0,1,0,1,0,1,0)
sexo <- c("F","F","F","F","F","F","F","M","F","M","F","M")

datos <- data.frame(id, t, d, sexo)

# Ordenar por ID
library(dplyr)

datos <- datos %>% arrange(id)
```

Información de los tiempos hasta el evento de los pacientes con cáncer

```
alineacion_columnas <- rep("c", ncol(datos)) # Esto centra todas las columnas

kable(datos, format = "html", align = alineacion_columnas) %>%
  kable_styling(
    full_width = TRUE,
    bootstrap_options = c("striped", "hover", "condensed")
  )
```

id	t	d	sexo
1	37	1	F
2	680	0	M
3	250	1	F
4	250	0	F
5	377	1	F
6	840	1	F
7	847	0	M
8	532	0	M
9	670	1	F
10	468	1	F
11	432	0	F
12	215	0	F

- $t$ : tiempo de seguimiento (días).
- $d$ : indicador de evento (1 = falla, 0 = censura).
- $sexo$ : variable de estratificación.

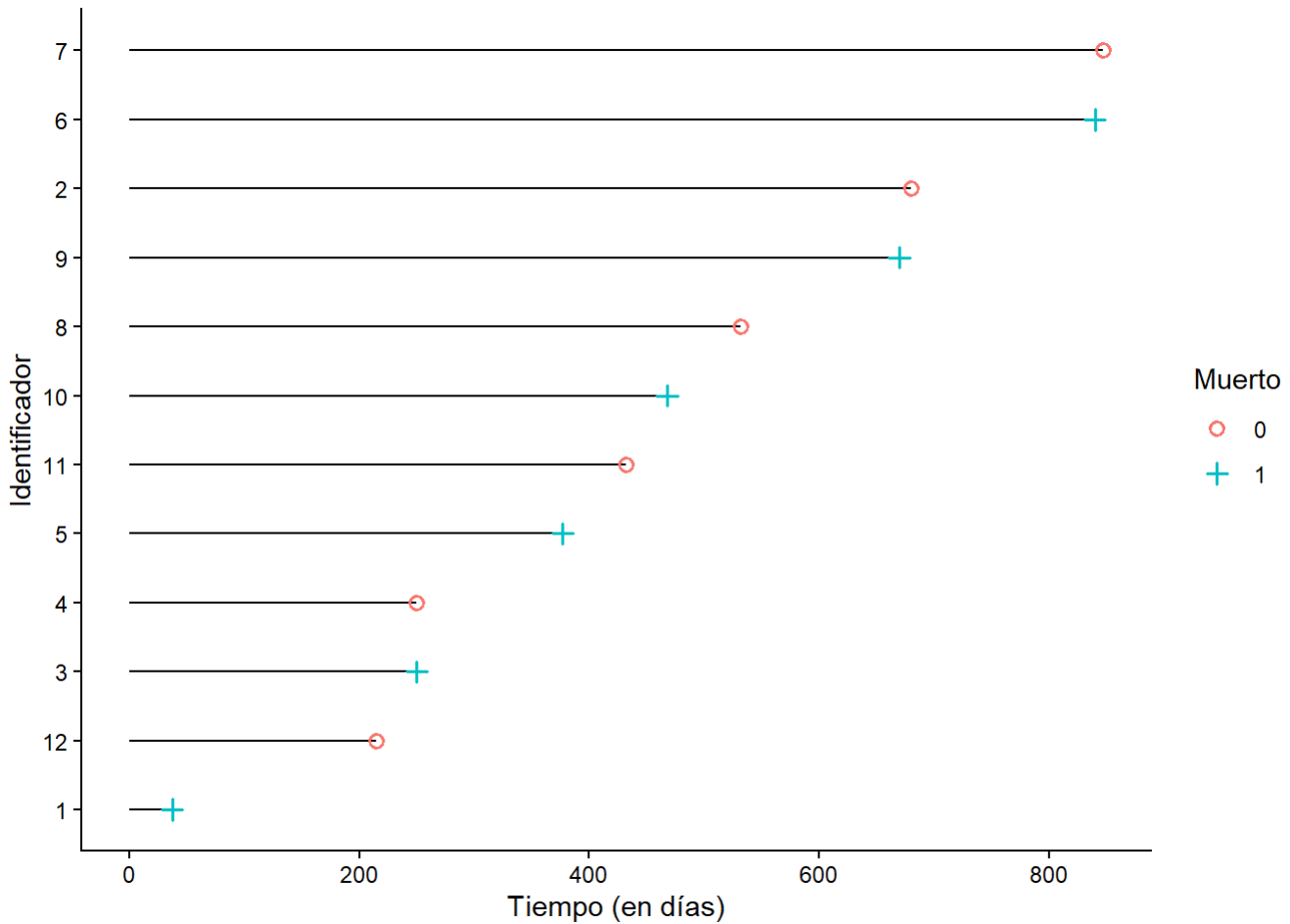
## Creando el gráfico de los tiempos de seguimiento

```
library(ggplot2)

library(dplyr)

datos <- datos %>%
  mutate(
    age_orig = 0,
    age_end = t,
    Muerto = factor(d, levels = c(0, 1)), # ← Cambio aquí
    id = factor(id, levels = id[order(age_end)])
  )

# Creando el gráfico
ggplot(datos, aes(x = id, y = age_end)) +
  geom_linerange(aes(ymin = age_orig, ymax = age_end)) +
  geom_point(aes(shape = Muerto, color = Muerto), stroke = 1, cex = 2) + # ← Y aquí
  scale_shape_manual(values = c(1, 3)) +
  guides(fill = FALSE) +
  labs(y = "Tiempo (en días)", x = "Identificador") +
  coord_flip() +
  theme_classic()
```



## Paso a paso Kaplan-Meier (cálculo manual)

La fórmula de Kaplan-Meier es:

$$\hat{S}(t) = \prod_{i:t_1 \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

donde:

- $t_i$ : tiempo de evento,
- $d_i$ : número de fallas en  $t_i$ ,
- $n_i$ : número en riesgo justo antes de  $t_i$ .

```
# Ordenamos por tiempo
datos_ord <- datos %>% arrange(t)

# Cálculo manual
tiempos_evento <- datos_ord %>% filter(d==1) %>% pull(t)
tabla_km <- data.frame(
  Tiempo = tiempos_evento,
  n_riesgo = NA,
  Fallas = NA,
  Censura = NA,
  Supervivencia = NA
)

n <- nrow(datos_ord)
S <- 1
for(i in seq_along(tiempos_evento)){
  t_i <- tiempos_evento[i]
  # en riesgo justo antes
  n_riesgo <- sum(datos_ord$t >= t_i)
  # fallas en t_i
  d_i <- sum(datos_ord$t==t_i & datos_ord$d==1)
  # censuras en t_i
  c_i <- sum(datos_ord$t==t_i & datos_ord$d==0)
  # prob
  S <- S * (1 - d_i/n_riesgo)

  tabla_km[i,] <- c(t_i, n_riesgo, d_i, c_i, round(S,4))
}
```

Tabla Kaplan-Meier para datos con censuras

Tiempo	En riesgo $n_i$	Fallas $d_i$	Censuras	Supervivencia $\hat{S}(t)$
37	12	1	0	$1 - 1/12 = 0.9167$
250	10	1	1	$0.9167 \times (1 - 1/10) = 0.825$
377	8	1	0	$0.825 \times (1 - 1/8) = 0.722$
468	7	1	0	$0.722 \times (1 - 1/7) = 0.619$
670	5	1	0	$0.619 \times (1 - 1/5) = 0.495$
840	3	1	0	$0.495 \times (1 - 1/3) = 0.330$

Realizando la curva de Kaplan Meier para datos censurados en R

Inicialmente creamos la base de datos con la información, para ello, creamos los vectores (usaremos la notación estándar)

```
id <- c(1, 12, 4, 3, 5, 11, 10, 8, 9, 2, 6, 7)
t <- c(37, 215, 250, 250, 377, 432, 468, 532, 670, 680, 840, 847)
d <- c(1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0)
sexo <- c("F", "F", "F", "F", "F", "F", "F", "M", "F", "M", "F", "M")
```

Posteriormente, creamos la base de datos

```
datos <- data.frame(id, t, d, sexo)
```

Nuevamente usamos las librerías “survival” y “survminer

La estructura es la siguiente:

```
library(survival)
library(survminer)

### Kaplan-Meier con survival para comparación
km <- survfit(Surv(t, d) ~ 1, data = datos)
```

Para ver la tabla:

```
tabla <- data.frame(
  Tiempo = km$time,
  En_Riesgo = km$n.risk,
  Fallas = km$n.event,
  Supervivencia = round(km$surv, 3)
)

knitr::kable(tabla, caption = "Tabla de Kaplan-Meier sin censuras")
```

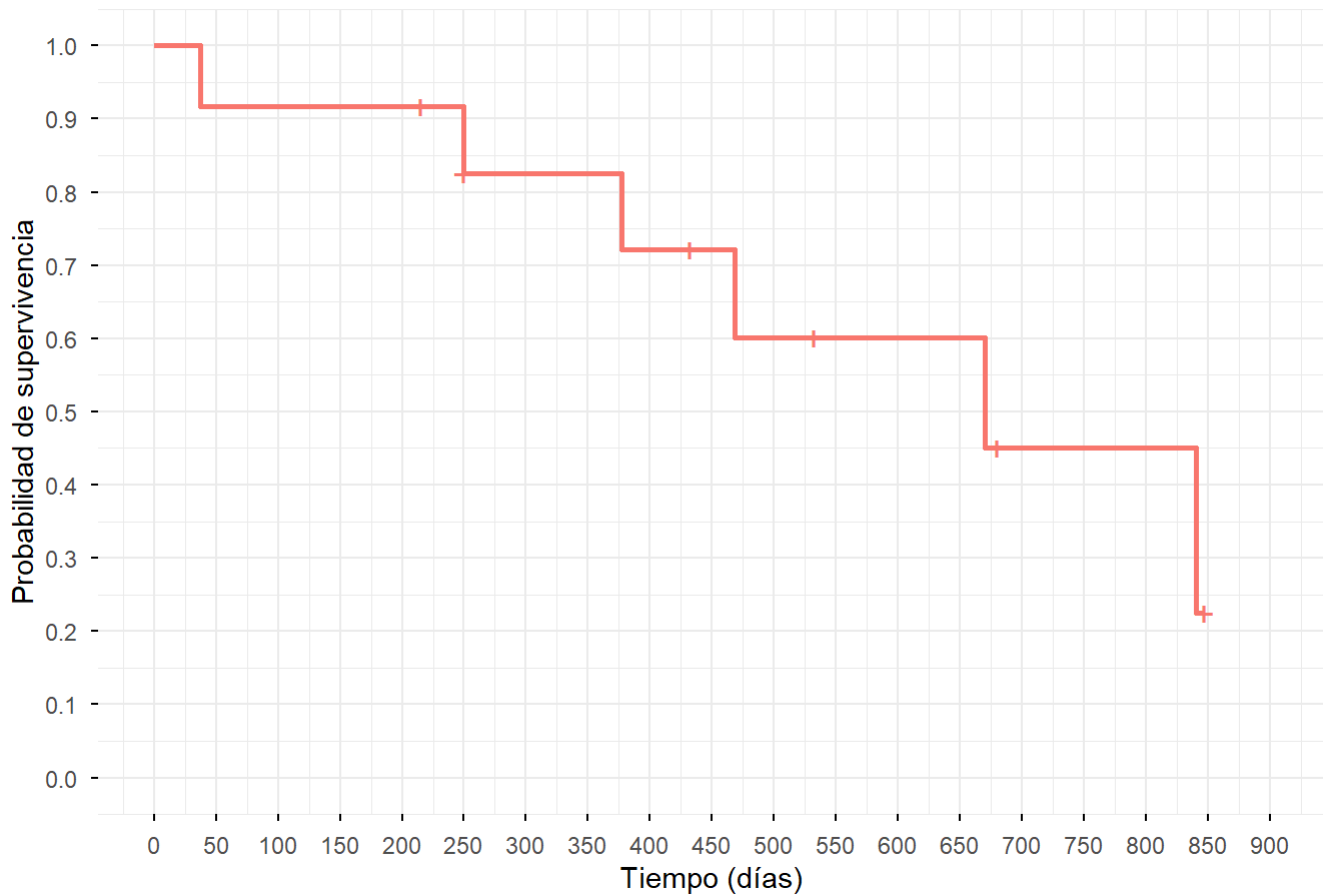
Tabla de Kaplan-Meier sin censuras

Tiempo	En_Riesgo	Fallas	Supervivencia
37	12	1	0.917
215	11	0	0.917
250	10	1	0.825
377	8	1	0.722
432	7	0	0.722
468	6	1	0.602
532	5	0	0.602
670	4	1	0.451
680	3	0	0.451
840	2	1	0.226
847	1	0	0.226

Para realizar el gráfico:

```
ggsurvplot(km,  
  conf.int = FALSE,  
  break.time.by = 50,  
  break.y.by = 0.1,  
  xlim = c(0, 900),  
  ggtheme = theme_minimal(),  
  xlab = "Tiempo (días)",  
  ylab = "Probabilidad de supervivencia",  
  title = "Gráfico de Kaplan-Meier sin censuras",  
  legend = "none")
```

Gráfico de Kaplan-Meier sin censuras



## Medidas Resumen de Supervivencia

Las medidas resumen de supervivencia son estadísticas utilizadas para describir y comparar la supervivencia entre grupos o tratamientos. Estas medidas proporcionan información sobre la duración típica de supervivencia, la probabilidad de supervivencia en un tiempo específico o el tiempo hasta que ocurre un evento en una proporción determinada de la población.

En el análisis de supervivencia, estas medidas se derivan generalmente de la curva de Kaplan-Meier. Algunas de las medidas resumen más utilizadas incluyen:

### Mediana de supervivencia

La mediana de supervivencia corresponde al tiempo en el cual el **50% de los individuos han experimentado el evento de interés**. Es una medida robusta y ampliamente utilizada en análisis de supervivencia porque no se ve afectada por valores extremos y puede estimarse incluso cuando existe censura.

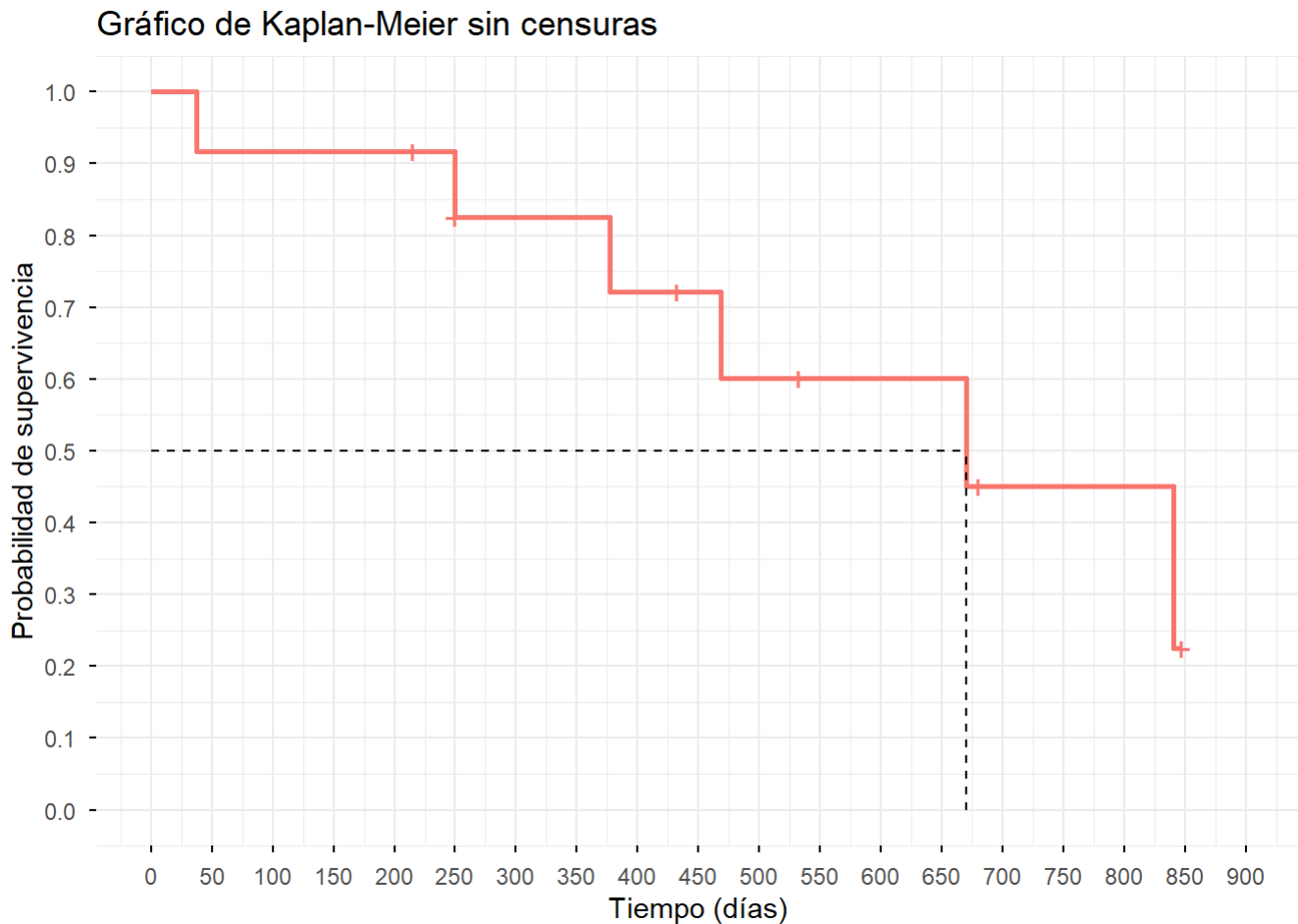
Ejemplo:

en R:

```
# Mediana de supervivencia
mediana <- summary(km)$table["median"]
mediana
## median
##      670
```

En el caso de nuestro ejemplo, la mediana es de 670 días, lo cual significa que el 50% de los pacientes han presentado el evento antes de los 670 días (que equivaldría a 1.83 años), mientras el otro 50% ha sobrevivido más allá de ese tiempo. Gráficamente, la mediana puede identificarse en la curva de supervivencia como el punto donde la curva cruza el valor **0.5** en el eje de supervivencia.

Veámoslo gráficamente:



Es recomendable reportar la mediana junto con su intervalo de confianza del 95%, ya que esto proporciona información sobre la precisión de la estimación.

## Percentiles de supervivencia

Los percentiles de supervivencia representan el tiempo en el cual una determinada proporción de individuos ha experimentado el evento de interés. En el contexto del análisis de supervivencia, estos percentiles se derivan de la función de supervivencia  $S(t)$ .

Formalmente, el percentil  $p$  del tiempo al evento corresponde al valor  $t_p$  que satisface:

$$S(t_p) = 1 - p$$

donde  $p$  representa la proporción de individuos que han experimentado el evento.

Por ejemplo:

- **Percentil 25 (Q1):** tiempo en el cual el 25% de los individuos han experimentado el evento ( $S(t) = 0.75$ ).
- **Percentil 75 (Q3):** tiempo en el cual el 75% de los individuos han experimentado el evento ( $S(t) = 0.25$ ).

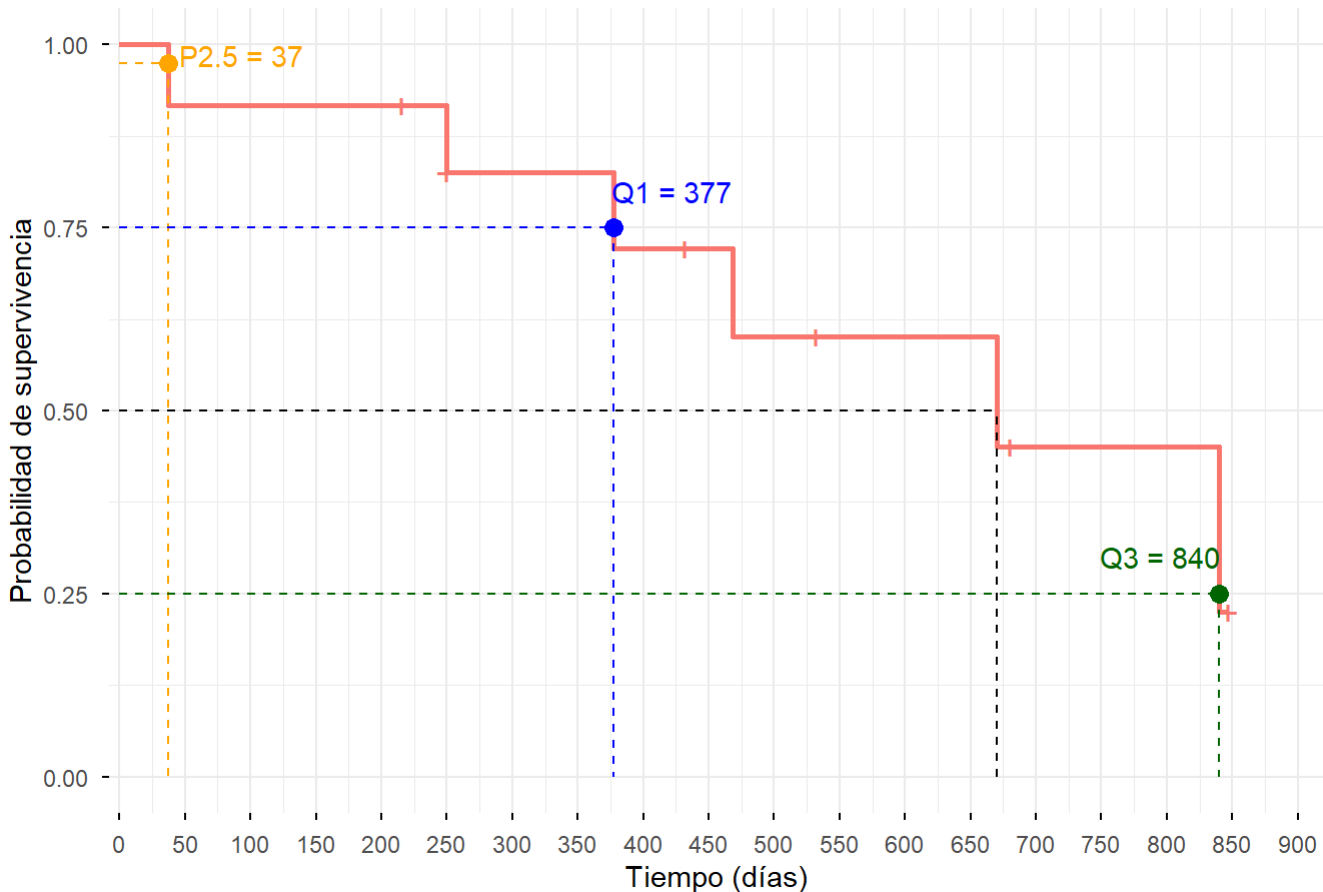
Estos percentiles proporcionan una descripción útil de la distribución de los tiempos al evento y pueden estimarse directamente a partir de la curva de Kaplan–Meier.

en R:

```
# Percentiles de supervivencia
percentiles <- quantile(km2, probs = c(0.025, 0.25, 0.75, 0.975))
percentiles$quantile
## 2.5 25.0 75.0 97.5
## 37 377 840 NA
```

Visualmente:

Curva de Kaplan-Meier



En este caso, el percentil 25 fue de 377 días (aproximadamente 1 año), lo que indica que el 25% de los pacientes han experimentado el evento antes de ese tiempo, o lo que es equivalente, que el 75% de los pacientes sobrevivió más de 1 año de seguimiento.

El percentil 75 fue de 840 días (aproximadamente 2.3 años), lo que indica que un 25% de los pacientes sobrevivió más de 2.3 años a partir del seguimiento.

Al evaluar los percentiles 2.5 y 97.5, encontramos que percentil 2.5 se encuentra alrededor de los 37 días, lo cual indica que el 2.5% de los pacientes han experimentado el evento antes de ese tiempo. Como se puede ver, el percentil 97.5 tampoco se pudo calcular, esto debido a que hay datos censurados y la curva de supervivencia no llega ese punto. Con la información disponible, lo que podemos afirmar, es que un 25% de los pacientes sobrevivió más de 840 días (es decir, más de 2.3 años).

## Intervalo de Confianza para la curva de supervivencia:

Los percentiles 2.5 y 97.5 no corresponden a los límites del intervalo de confianza del 95%. Estos percentiles describen únicamente puntos específicos de la distribución del tiempo al evento y no deben confundirse con medidas de incertidumbre estadística. En el caso de la mediana de supervivencia, el intervalo de confianza del 95% se obtiene a partir de la varianza de Greenwood, la cual cuantifica la variabilidad del estimador de Kaplan–Meier de la función de supervivencia. Además, para mejorar el comportamiento del intervalo –particularmente cuando las probabilidades de supervivencia se aproximan a 0 o 1– suele aplicarse una transformación log–log, que permite mantener los límites del intervalo dentro del rango válido de probabilidades.

Veamos como lo calculamos en R:

```
surv_median(km)
##   strata median lower upper
## 1     All    670   377   NA
```

Con un nivel de confianza del 95%, se puede afirmar que la mediana poblacional de supervivencia (en esta base de datos sin censura) es al menos de 377 días. El límite superior del intervalo de confianza no pudo estimarse, ya que la información proporcionada por la curva de supervivencia no es suficiente para identificar ese extremo del intervalo.

## Media Restringida:

La media restringida de supervivencia (RMST, por sus siglas en inglés) es una medida que representa el tiempo promedio libre del evento hasta un horizonte temporal específico  $\tau$ . Se calcula como el área bajo la curva de supervivencia desde el tiempo 0 hasta  $\tau$ :

$$RMST(\tau) = \int_0^{\tau} S(t) dt$$

Esta medida es especialmente útil cuando la curva de supervivencia no llega a 0, lo que puede ocurrir en estudios con censura o con eventos raros. La RMST además proporciona una estimación del tiempo promedio de supervivencia dentro de un período de tiempo definido, lo que puede ser más informativo que la mediana en ciertos contextos.

### Para el tiempo total de seguimiento tenemos:

```
summary(km2)$table
```

##	records	n.max	n.start	events	rmean	se(rmean)	median	0.95LCL
##	12.0000	12.0000	12.0000	6.0000	602.5096	83.9467	670.0000	377.0000
##	0.95UCL							
##	NA							

El tiempo medio restringido de supervivencia (RMST) fue de 466.5 días (error estándar = 71.0 días), lo que representa el tiempo promedio libre del evento durante el **periodo de observación**.

También podemos calcular la media restringida para un período específico, como 1 año (365 días) o 2 años (730 días), así:

### Para el primer año tenemos:

```
summary(km2, rmean = 365)$table
```

##	records	n.max	n.start	events	rmean	se(rmean)	median	0.95LCL
##	12.00000	12.00000	12.00000	6.00000	327.12500	27.16035	670.00000	377.00000
##	0.95UCL							
##	NA							

### Para el segundo año tenemos:

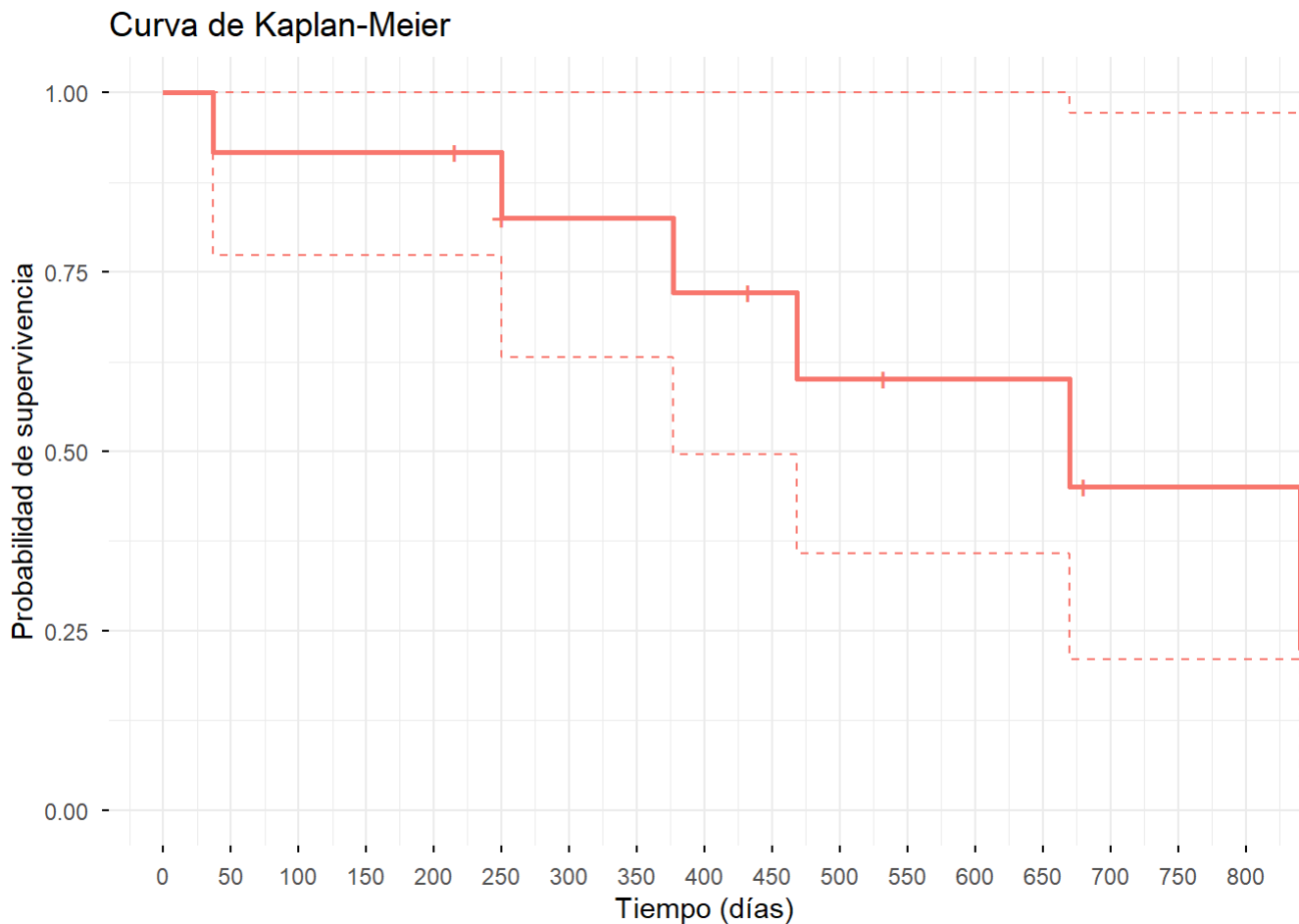
```
summary(km2, rmean = 730)$table
```

##	records	n.max	n.start	events	rmean	se(rmean)	median	0.95LCL
##	12.00000	12.00000	12.00000	6.00000	551.30156	68.93309	670.00000	377.00000
##	0.95UCL							
##	NA							

A partir de estos resultados se puede interpretar que, durante el primer año de seguimiento, los pacientes permanecieron libres del evento en promedio 306 días (error estándar = 28.23). Cuando el horizonte de restricción se amplía a dos años, el tiempo medio restringido de supervivencia

indica que los pacientes permanecieron libres del evento en promedio 466.5 días dentro de ese período de seguimiento.

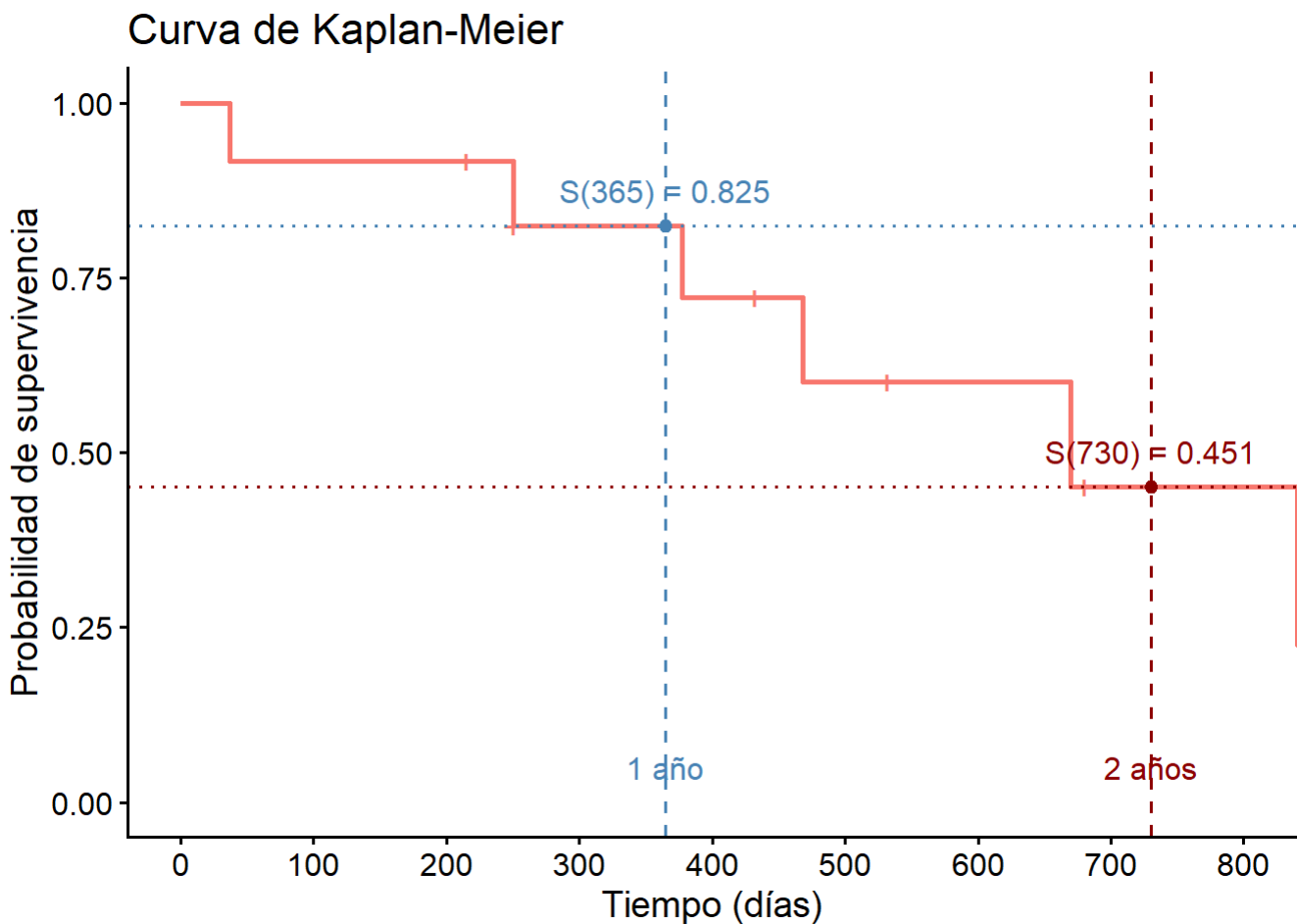
Para el caso del ejemplo, hacemos el grafico con los intervalos:



## Supervivencia a un tiempo específico

Corresponde a la probabilidad de que un individuo sobreviva más allá de un tiempo específico (por ejemplo, 1 año, 5 años, etc.). Para determinar esta medida, se puede observar la curva de supervivencia y encontrar el valor en el eje de supervivencia correspondiente al tiempo específico de interés. El valor en el eje de supervivencia en ese punto indicará la probabilidad de supervivencia a ese tiempo específico.

Para el caso del ejemplo, ubicamos los años en el siguiente gráfico:



Para un tiempo específico de 1 año (365 días), podemos observar la curva de supervivencia y encontrar el valor en el eje de supervivencia correspondiente a los 365 días. Para dos años de seguimiento, podemos decir que el 45% de los pacientes sobreviven más allá de ese tiempo, lo que se puede observar en la curva de supervivencia donde el valor en el eje de supervivencia es 0.45 en el punto correspondiente a los 2 años.

## Comparación de Curvas de Supervivencia

Para comparar las curvas de supervivencia entre dos o más grupos, se pueden utilizar pruebas estadísticas como la prueba de log-rank, que evalúa si hay diferencias significativas en la supervivencia entre los grupos a lo largo del tiempo. Además, se pueden calcular razones de riesgos (hazard ratios) para cuantificar la magnitud de las diferencias en la supervivencia entre los grupos.

*Prueba de Log rank.* Es una prueba no paramétrica que compara las curvas de supervivencia entre dos o más grupos. Evalúa si hay diferencias significativas en la supervivencia a lo largo del tiempo, considerando el número de eventos y el número de individuos en riesgo en cada grupo.

*Las Hipótesis a probar son las siguientes:*

- La hipótesis nula de la prueba es que no hay diferencias entre las curvas de supervivencia entre los grupos
- La hipótesis alterna es que son diferentes, es decir, al menos un grupo tiene una curva de supervivencia diferente.

*Estadístico de la prueba de log-rank:*

La fórmula para la prueba de log-rank es:

$$\chi^2 = \frac{(O_1 - E_1)^2}{V_1}$$

donde:

- **$O_1$** : número observado de eventos en el grupo 1,
- **$E_1$** : número esperado de eventos en el grupo 1 bajo la hipótesis nula de que no hay diferencias en la supervivencia entre los grupos,
- **$V_1$** : varianza del número de eventos en el grupo 1.

Como podemos ver, se trata de una prueba chi-cuadrado para muestras grandes e independientes, por lo que el estadístico de la prueba se distribuye aproximadamente como una distribución chi-cuadrado con grados de libertad igual al número de grupos menos uno, por lo tanto, se compara



## Prueba de *Log-rank* para comparar las curvas de supervivencia entre los grupos

```
survdif(Surv(t,d) ~ sexo, data=datos)

## Call:
## survdiff(formula = Surv(t, d) ~ sexo, data = datos)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## sexo=F  9         6      3.58      1.64      4.26
## sexo=M  3         0      2.42      2.42      4.26
##
## Chisq= 4.3 on 1 degrees of freedom, p= 0.04
```

El resultado de la prueba de log-rank nos indica que hay diferencias estadísticamente significativas entre las curvas de supervivencia entre el grupo de hombres y el grupo de mujeres ( $p=0.039$ ). Podemos concluir, que en esta muestra de pacientes, los hombres presentan una mejor supervivencia que las mujeres, lo cual se puede observar en la curva de supervivencia por sexo, donde la curva correspondiente a los hombres se encuentra por encima de la curva correspondiente a las mujeres. Sin embargo, hay que ser muy cautelosos al interpretar estos resultados, ya que se trata de una muestra pequeña y no se han ajustado por posibles factores de confusión que podrían estar influyendo en la supervivencia, como la edad, el estadio del cáncer, el tipo de tratamiento recibido, entre otros.

## Lecturas complementarias:

- **Survival analysis: A self-learning text.** David G. Kleinbaum and Mitchel Klein. Springer, New York, 2012.
- **Función de riesgo instantáneo:** Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.
- **Función de riesgo acumulado:** Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4), 945-966.
- **R:** Therneau, T. M. (2026). A package for survival analysis in R. R package version 3.8-6, <https://CRAN.R-project.org/package=survival>.
- klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.