

# Dos nuevas librerías para análisis de la supervivencia: dcens y bwsurvival

Carles Serrat\* (UPC)  
carles.serrat@upc.edu

Olga Julià\* (UB), Jorge Cortés (UPC-ICO), Victoria Moneta (UPC-PSSJD) y  
Guadalupe Gómez\* (UPC)

Grup de Recerca en Anàlisi eStadística de la Supervivència (GRASS)

\* Miembros de la Red BIOSTATNET

III Jornada de Usuarios de R  
Madrid, 17 y 18 de Noviembre de 2011

## Guión de la presentación

- Librería dcens
- Librería bwsurvival
- Futuras extensiones

## Datos con presencia de doble censura

**Fuente** Unidad de desintoxicación del Hospital Universitario 'Germans Trias i Pujol' en Badalona.

**Datos** 232 drogadictos interparentales entre 1985 y 1997 se infectaron con el virus del SIDA

## Datos con presencia de doble censura

**Fuente** Unidad de desintoxicación del Hospital Universitario 'Germans Trias i Pujol' en Badalona.

**Datos** 232 drogadictos interparentales entre 1985 y 1997 se infectaron con el virus del SIDA

**Objetivo** Estudiar el tiempo desde el inicio del consumo de drogas intravenosas y el diagnóstico del SIDA.

## Datos con presencia de doble censura

**Fuente** Unidad de desintoxicación del Hospital Universitario 'Germans Trias i Pujol' en Badalona.

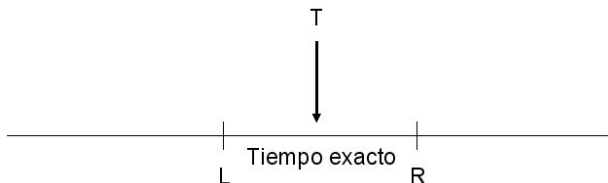
**Datos** 232 drogadictos interparentales entre 1985 y 1997 se infectaron con el virus del SIDA

**Objetivo** Estudiar el tiempo desde el inicio del consumo de drogas intravenosas y el diagnóstico del SIDA.

**Doble Censura** Este lapso de tiempo no se observa siempre de forma completa. Hay pacientes que en el momento de cerrar el estudio aún siguen libres de SIDA (datos censurados por la derecha), mientras que otros mueren de SIDA sin diagnóstico previo (datos censurados por la izquierda).

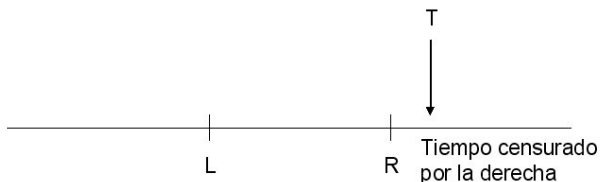
## Esquema de la doble censura

$T$  es una variable positiva (tiempo hasta el evento de interés)  
 $0 \leq L \leq R$  c.s. independientes de  $T$



## Esquema de la doble censura

$T$  es una variable positiva (tiempo hasta el evento de interés)  
 $0 \leq L \leq R$  c.s. independientes de  $T$



## Esquema de la doble censura

$T$  es una variable positiva (tiempo hasta el evento de interés)  
 $0 \leq L \leq R$  c.s. independientes de  $T$





# Notación

Una muestra de tamaño  $n$  está formada por

$$\{(T_i, L_i, R_i) \quad i = 1 \dots, n\},$$

pero sólo observamos  $\{(U_i, \delta_i) \quad i = 1 \dots n\}$  donde

$$U_i = (L_i \vee T_i) \wedge R_i \quad \text{y}$$

$$\delta_i = \mathbf{1}_{\{T_i > R_i\}} - \mathbf{1}_{\{T_i < L_i\}} = \begin{cases} -1 & \text{censura izquierda} \\ 0 & \text{exacto} \\ 1 & \text{censura derecha} \end{cases} .$$

## Notación

Una muestra de tamaño  $n$  está formada por

$$\{(T_i, L_i, R_i) \quad i = 1 \dots, n\},$$

pero sólo observamos  $\{(U_i, \delta_i) \quad i = 1 \dots n\}$  donde

$$U_i = (L_i \vee T_i) \wedge R_i \quad \text{y}$$

$$\delta_i = \mathbf{1}_{\{T_i > R_i\}} - \mathbf{1}_{\{T_i < L_i\}} = \begin{cases} -1 & \text{censura izquierda} \\ 0 & \text{exacto} \\ 1 & \text{censura derecha} \end{cases} .$$

- Julià, O. y Gómez, G. (2011) Simultaneous marginal survival estimators when doubly censored data is present. *Lifetime Data Analysis* **17**:347-372.

## Estimación Simultánea de $S_T$ , $S_L$ y $S_R$

Si denotamos por  $o_j$ ,  $j = 1 \dots r$  los distintos tiempos ordenados a partir de  $u_{(1)} \leq u_{(2)} \leq \dots u_{(n)}$  y admitiendo empates,

$$\left. \begin{aligned}
 \hat{S}_T(t) &= \begin{cases} 1 & \forall 0 < t < o_1 \\ \frac{1}{n} \sum_{j; o_j > t} \frac{d_j}{\hat{S}_R(o_j^-) - \hat{S}_L(o_j)} & \forall t \geq o_1 \end{cases} \\
 \hat{S}_L(t) &= \frac{1}{n} \sum_{j; o_j > t} \frac{\mu_j}{1 - \hat{S}_T(o_j^-)} & \forall t > 0 \\
 \hat{S}_R(t) &= 1 - \frac{1}{n} \sum_{j; o_j \leq t} \frac{\lambda_j}{\hat{S}_T(o_j)} & \forall t > 0.
 \end{aligned} \right\}$$

donde  $d_j$ ,  $\mu_j$  y  $\lambda_j$  es el número de observaciones exactas, censuradas por la izquierda y censuradas por la derecha en  $o_j$ , respectivamente.

# Inverse Weighted Estimators $\hat{S}_T$ , $\hat{S}_L$ and $\hat{S}_R$

Los estimadores propuestos para  $S_T(t)$ ,  $S_L(t)$  y  $S_R(t)$  son la solución del sistema de ecuaciones anterior.

$\hat{S}_T(t)$ ,  $\hat{S}_L(t)$  y  $\hat{S}_R(t)$  cumplen:

- 1 toman valores entre 0 y 1
- 2 son funciones a saltos continuas por la derecha
- 3 saltan únicamente en los puntos  $o_j$
- 4 se pueden expresar por

$$\hat{S}_T(t) = \sum_{j;o_j > t} f_j, \quad \hat{S}_L(t) = \sum_{j;o_j > t} h_j, \quad \hat{S}_R(t) = 1 - \sum_{j;o_j \leq t} g_j,$$

donde  $f_j$ ,  $h_j$  y  $g_j$  son las medidas de los saltos en  $o_j$ .

# Librería dcens

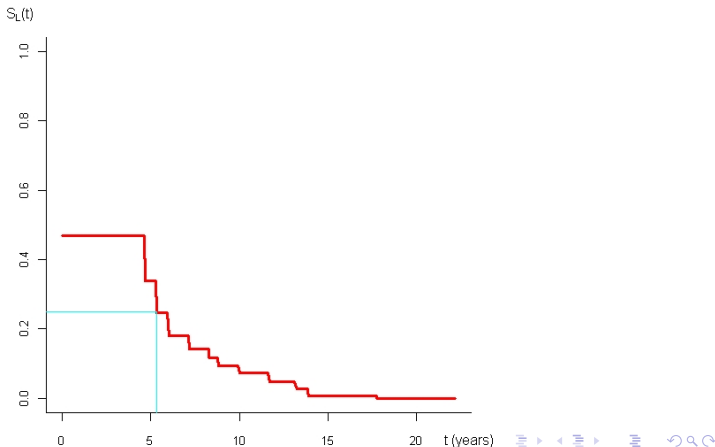
This R package implements the doubly censoring estimation methodology described at 'Simultaneous marginal survival estimators when doubly censored data is present' (O. Julià and G. Gómez, 2011).

## Functions

- `dcens` estimates quantiles and bootstrap standard errors for  $S_T$ ,  $S_L$  and  $S_R$  survival functions.
- `plot.dcens` plots the estimated  $S_T$ ,  $S_L$  and  $S_R$  functions.
- `survivalSL` estimates the  $S_L$  survival function.
- `survivalST` estimates the  $S_T$  survival function.
- `survivalSR` estimates the  $S_R$  survival function.
- `survQuantile` computes the quantiles from the estimated  $S_T$ ,  $S_L$  and  $S_R$  survival functions.

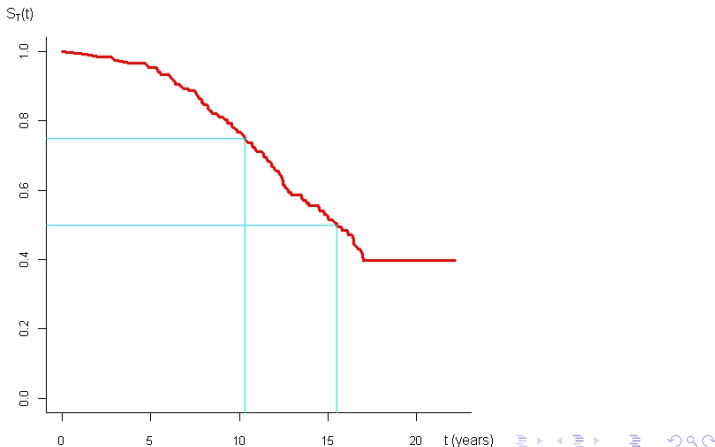
# Resultados gráficos

survivalSL



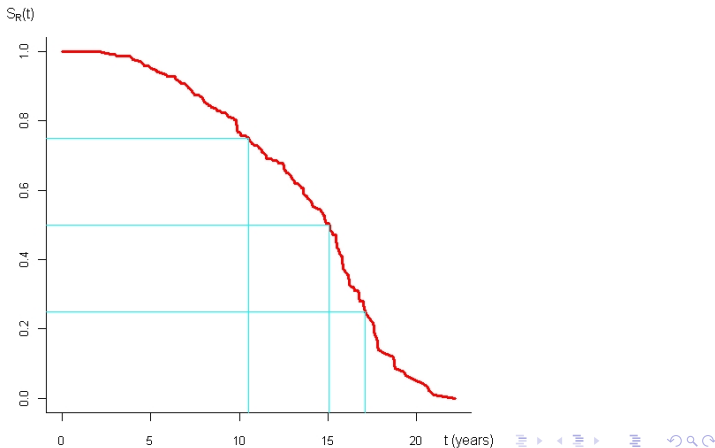
# Resultados gráficos

survivalST



# Resultados gráficos

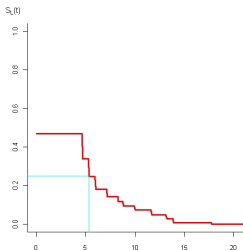
survivalSR



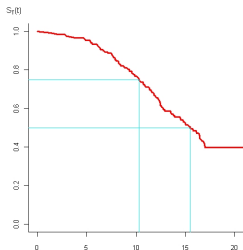


# Resultados gráficos

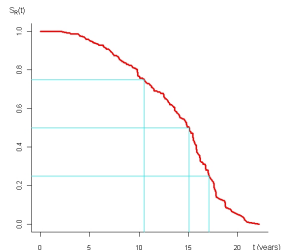
survivalSL



survivalST



survivalSR



## Conjunto de datos que motiva el análisis

### La cohorte de pacientes esquizofrénicos del PSSJD

- HCl de pacientes con diagnóstico de esquizofrenia cuya primera visita a algunos de los centros de la red del PSSJD-SSM.
- Inicio del estudio: 01/01/2002.
- Fin del estudio: 31/03/2008.

## Conjunto de datos que motiva el análisis

### La cohorte de pacientes esquizofrénicos del PSSJD

- HCI de pacientes con diagnóstico de esquizofrenia cuya primera visita a algunos de los centros de la red del PSSJD-SSM.
- Inicio del estudio: 01/01/2002.
- Fin del estudio: 31/03/2008.

### Objetivo científico

- Estudio de la efectividad de un segundo tratamiento condicionado a la efectividad de un primer tratamiento.

## Descripción de la base de datos

- *ID*: ID del paciente.
- *Fvis1*: Fecha de la primera visita del paciente (entrada al estudio del paciente).
- *Fvis2*: Fecha de la segunda visita (primer cambio de medicación desde la entrada al estudio).
- *Fvisj*: Fecha de la *j*-ésima visita con cambio de medicación.
- *Medicación 1*: Fármaco/s prescrito/s en el momento basal.
- *Medicación 2*: Fármaco/s prescrito/s en la segunda visita.
- *Medicación j*: Fármaco/s prescrito/s en la *j*-ésima visita.
- $T_1$ : Número de días que el paciente mantiene la medicación 1.
- $T_2$ : Número de días de mantenimiento de la segunda medicación (desde la segunda a la tercera visita).
- $T_j$ : Número de días de mantenimiento de la *j*-ésima medicación.
- *C*: Seguimiento total del paciente (en días).
- *Sexo*: Sexo del paciente
- *Edad*: Edad del paciente al inicio del estudio.
- *Ecivil*: Estado civil del paciente.
- *Diag*: Diagnóstico del paciente.



# Notación

- $T_1$  y  $T_2$ : Dos tiempos sucesivos a eventos  $E_1$  y  $E_2$ , respectivamente.
- $C$ : Tiempo de seguimiento.
- Se asume que  $T_1$ ,  $T_2$  y  $T_1+T_2$  son independientes de  $C$ .
- $T_1$  está censurado por la derecha por  $C$
- $T_2$  está censurado por la derecha por  $C-T_1$

# Notación

- $T_1$  y  $T_2$ : Dos tiempos sucesivos a eventos  $E_1$  y  $E_2$ , respectivamente.
- $C$ : Tiempo de seguimiento.
- Se asume que  $T_1$ ,  $T_2$  y  $T_1+T_2$  son independientes de  $C$ .
- $T_1$  está censurado por la derecha por  $C$
- $T_2$  está censurado por la derecha por  $C-T_1$

→  $T_2$  está sujeto a censura dependiente

# Notación

- $T_1$  y  $T_2$ : Dos tiempos sucesivos a eventos  $E_1$  y  $E_2$ , respectivamente.
- $C$ : Tiempo de seguimiento.
- Se asume que  $T_1$ ,  $T_2$  y  $T_1+T_2$  son independientes de  $C$ .
- $T_1$  está censurado por la derecha por  $C$
- $T_2$  está censurado por la derecha por  $C-T_1$

→  $T_2$  está sujeto a censura dependiente

Para estimar  $T_2$  no se pueden utilizar los métodos clásicos de supervivencia para censura independiente. Para ello utilizamos la metodología propuesta por **Serrat, C. y Gómez, G. (2007)**.

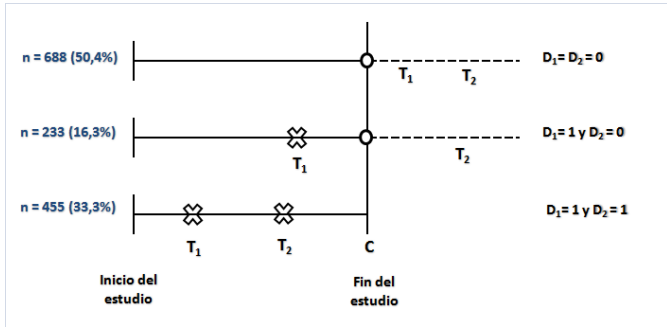


# Notación

¿De donde partimos?

$(Y_1, Y_2, D_1, D_2)$ ,  $j=1,2$ ,  $Y_j = \min\{T_j, C_j\}$ ,  $D_j = 1\{T_j \leq C_j\}$ ,  $C_1 = C$ ,  
 $C_2 = (C - T_1)1\{T_1 \leq C\}$

¿Qué situaciones se nos presentan?



# El método de Gómez y Serrat

## Los autores plantean:

- $0 < \tau_1 < \tau_2 < \dots < \tau_M$  los  $M$  tiempos de interés
- $T_1$  es particionado en  $M + 1$  categorías:  
 $I_k = (\tau_{k-1}, \tau_k]$ ,  $k = 1, \dots, M$ ,  $I_{M+1} = (\tau_M, \tau_{M+1})$
- $T_2$  sólo es observado si  $T_1 + T_2$  no excede el tiempo de seguimiento total.
- La probabilidad de estar a riesgo en el momento  $b$  para  $T_2$ , para un individuo cuyo  $T_1 = t_{1i}$ , es igual a  
$$\text{Prob}\{T_2 \geq b\} \text{Prob}(C > t_{1i} + b) = G(t_{1i} + b).$$

# El método de Gómez y Serrat

## Función de riesgo condicionada:

$$\hat{\lambda}_{T_2|T_1 \in I_k}^{\hat{G}}(b) = \frac{\sum_{i=1}^n \mathbf{1}\{Y_{1i} \in I_k, D_{1i} = 1, Y_{2i} = b, D_{2i} = 1\} / \hat{G}(Y_{1i} + b)}{\sum_{i=1}^n \mathbf{1}\{Y_{1i} \in I_k, D_{1i} = 1, Y_{2i} \geq b\} / \hat{G}(Y_{1i} + b)} \quad (1)$$

## Función de supervivencia condicionada:

$$\hat{S}_{T_2|T_1 \in I_k}^{\hat{G}}(t) = \prod_{b \leq t} \{1 - \hat{\lambda}_{T_2|T_1 \in I_k}^{\hat{G}}(b)\} \quad (2)$$

# Librería bwsurvival

This R package implements the conditional survival function described in *Nonparametric bivariate estimation for successive survival times* (Serrat and Gómez, 2007).

## Functions

- `bw21`: Conditional survival for  $T_2$  given categories of  $T_1$ .
- `bw21x`: Conditional survival for  $T_2$  given categories of  $T_1$  adjusted for  $x$ .
- `plot21`: Plot for Conditional survival for  $T_2$  given categories of  $T_1$ .
- `plot21x`: Plot for Conditional survival for  $T_2$  given categories of  $T_1$  adjusted for  $x$ .
- `plot.bwsurvival`: Plot for bwsurvival object.
- `summary.bwsurvival`: Conditional survival for times of interest of  $T_2$  given categories of  $T_1$ .

# Librería bwsurvival

## Package 'bwsurvival'

November 7, 2011

**Type** Package

**Title** Nonparametric conditional survival function for successive survival times

**Version** 1.0

**Date** 2011-11-01

**Author** C. Serrat, G. Gómez, V. Moneta

**Maintainer** Victoria Moneta <m victoria.moneta@pssjd.org> and Carles Serrat <carles.serrat@upc.edu>

**Description** This R package implements the conditional survival function described at 'Nonparametric bivariate estimation for successive survival times' (C. Serrat and G. Gomez, 2007).

**Depends** survival

# Librería bwsurvival

```
R Console
Archivo  Editor  Misc  Paquetes  Ventanas  Ayuda

Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

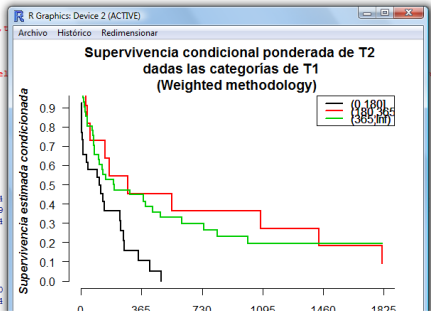
> library(bwsurvival)
Loading required package: survival
Loading required package: splines
> data(times)
> bw21_farm<-bw21(times$days2,times$scens2,times$days1,t
> #Funcion de supervivencia condicionada#
> x11()# abro una ventana gráfica#
> par(font=1,font.axis=1,font.lab=4,las=1)
> plot21(bw21_farm,xlab='Dias hasta la interrupción de
> axis(1,at=seq(0,2555,365))
> axis(2,at=seq(0,1,.1))
> bw21_farm
$breaks
[1] 180 365

$counts
[1] 31 11 57

$times1
[1] 1 3 5 12 15 17 28 35 42 54
[27] 231 242 245 276 291 300 329 368 385 389
[53] 701 754 756 766 833 882 888 900 987 1204

$K1
[1] 71

$times2
[1] 1 4 6 13 14 16 21 24 28 30
[27] 128 131 133 139 144 151 170 195 202 234
```



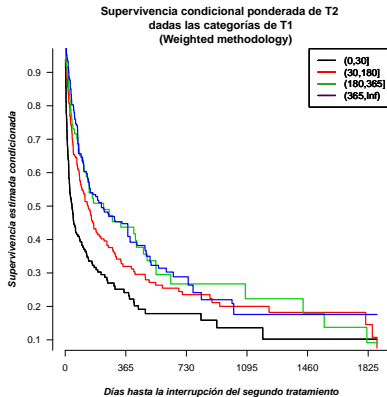
# Estimación no paramétrica de la eficiencia del segundo tratamiento condicionado al primero

`summary.bwsurvival`

Tiempos $T_2$	$T_1 \in (0,30]$	$T_1 \in (30,180]$	$T_1 \in (180,365]$	$T_1 \in (365,2248]$
7	0,7720	0,9009	0,9173	0,9546
15	0,6651	0,8572	0,8886	0,9361
30	0,5343	0,7664	0,8028	0,8683
45	0,4803	0,6902	0,7442	0,8031
60	0,4358	0,6497	0,7162	0,7592
75	0,4149	0,6121	0,6864	0,7103
90	0,3933	0,5761	0,6519	0,6571
120	0,3649	0,5133	0,6021	0,6035
150	0,3353	0,4849	0,5394	0,5529
180	0,3165	0,4321	0,5087	0,5343
240	0,2868	0,3956	0,4899	0,4824
300	0,2512	0,3561	0,4536	0,4531
365	0,2414	0,3196	0,4370	0,4475
540	0,1785	0,2716	0,3367	0,3228
730	0,1785	0,2352	0,2672	0,2885
1095	0,1362	0,2000	0,2229	0,1759
1460	0,1020	0,1820	0,1792	0,1759
1825	0,1020	0,1457	0,0914	0,1759
2190	0,1020	0,0751	0,0914	0,1759

# Estimación no paramétrica de la eficiencia del segundo tratamiento condicionado al primero

plot21

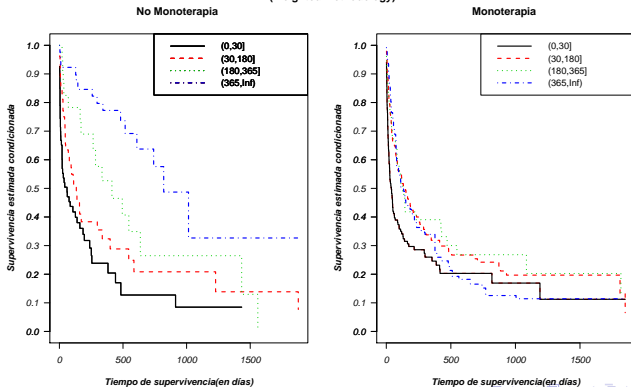




# Estimación no paramétrica de la eficiencia del segundo tratamiento condicionado al primero ajustado según tratamiento basal

plot21x

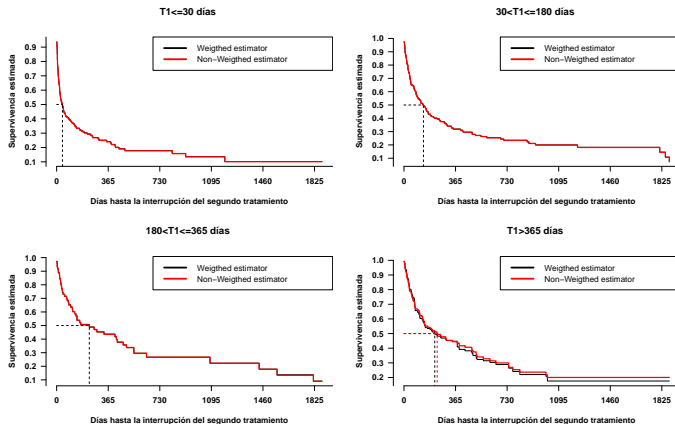
Supervivencia condicional estratificada de T2 dadas las categorías de T1 ajustado según tratamiento basal (Weighted methodology)



## Comparación Weighted vs Non Weighted Methodology

wmet=1 vs wmet=0

Función de supervivencia ponderada condicionada vs. Kaplan-Meier estratificado en cada una de las categorías de T1



# Futuras Extensiones

- Librería dcens
  - 1 Permitir para el cálculo de cuantiles alternativas a la interpolación lineal
  - 2 Incorporar la estimación de los errores estándar haciendo uso de la distribución asintótica
- Librería bwsurvival
  - 1 Incorporar la estimación de los errores estándar via bootstrap
  - 2 Incluir la posibilidad de un contraste de hipótesis de igualdad de distribuciones

# Muchas gracias por su atención