

# Análisis de supervivencia

Sesión 7. Curso: “Análisis estadístico aplicado con R Commander”

Carlos Vergara Hernández  
vergara\_car@gva.es

## Algunos conceptos clave

# Los datos censurados

La principal característica del análisis de supervivencia es la aparición de observaciones censuradas como consecuencia de:

# Los datos censurados

La principal característica del análisis de supervivencia es la aparición de observaciones censuradas como consecuencia de:

- una pérdida de seguimiento: una razón ajena al estudio impide completar el seguimiento.

# Los datos censurados

La principal característica del análisis de supervivencia es la aparición de observaciones censuradas como consecuencia de:

- una pérdida de seguimiento: una razón ajena al estudio impide completar el seguimiento.
- la exclusión del estudio: hay razones derivadas del estudio que hacen que el sujeto deje de participar en él

# Los datos censurados

La principal característica del análisis de supervivencia es la aparición de observaciones censuradas como consecuencia de:

- una pérdida de seguimiento: una razón ajena al estudio impide completar el seguimiento.
- la exclusión del estudio: hay razones derivadas del estudio que hacen que el sujeto deje de participar en él
- la finalización del estudio: el estudio finaliza y el sujeto no desarrolla el evento.

# Los datos censurados

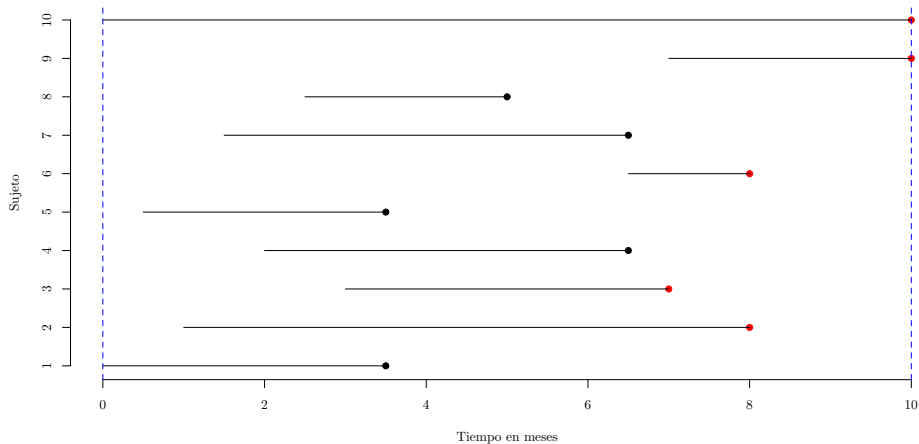
La principal característica del análisis de supervivencia es la aparición de observaciones censuradas como consecuencia de:

- una pérdida de seguimiento: una razón ajena al estudio impide completar el seguimiento.
- la exclusión del estudio: hay razones derivadas del estudio que hacen que el sujeto deje de participar en él
- la finalización del estudio: el estudio finaliza y el sujeto no desarrolla el evento.

## Varios tipos de censura

Hay varios tipos de censura: izquierda, derecha o doble, aunque aquí solo hablaremos de la censura a la derecha (que es la más frecuente).

# Ejemplo 1

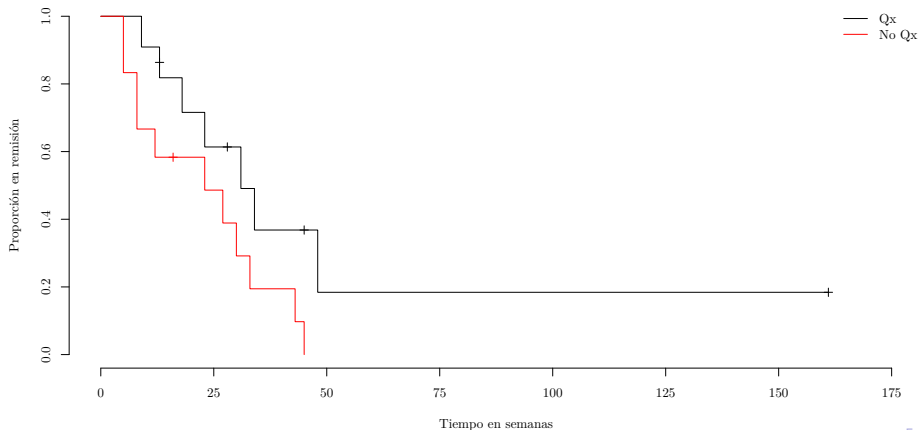




## Ejemplo 2

**Tabla 1.** Ejemplo de presentación de datos de supervivencia.

Grupo	Supervivencia
Maintained	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+
Nonmaintained	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45



## Si se obvia la censura...

- Se obtiene un resultado sesgado (valores más pequeños que los reales)
  - Las observaciones censuradas tienden a tener un tiempo de supervivencia más largo.

## Si se obvia la censura...

- Se obtiene un resultado sesgado (valores más pequeños que los reales)
  - Las observaciones censuradas tienden a tener un tiempo de supervivencia más largo.

	Todos los datos		Solo datos completos	
	Media	Mediana	Media	Mediana
Qx mantenida	38.45	28.0	25.14	23
Qx no mantenida	21.25	19.5	21.73	23

## Vocabulario: función de supervivencia

La función de supervivencia  $S(t)$  proporciona la probabilidad de que un sujeto sobreviva un determinado tiempo.

La función de supervivencia  $S(t)$  proporciona la probabilidad de que un sujeto sobreviva un determinado tiempo.

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

La función de supervivencia  $S(t)$  proporciona la probabilidad de que un sujeto sobreviva un determinado tiempo.

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

La función de supervivencia siempre mostrará un descenso al inicio por el hecho de haber más individuos expuestos al riesgo a medida que pasa el tiempo, y es por ello que los descensos al inicio o final no sean comparables.

## Vocabulario: función de riesgo

La función de riesgo es la proporción de casos que presentan el evento en un momento determinado sobre el número de casos que llegan a ese momento.

## Vocabulario: función de riesgo

La función de riesgo es la proporción de casos que presentan el evento en un momento determinado sobre el número de casos que llegan a ese momento.

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



# Vocabulario: función de riesgo

La función de riesgo es la proporción de casos que presentan el evento en un momento determinado sobre el número de casos que llegan a ese momento.

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

Las funciones de riesgo pueden presentar diferentes formas que responden a diferentes situaciones.

# Objetivos del análisis de supervivencia

Los objetivos básicos del análisis de supervivencia son tres:

# Objetivos del análisis de supervivencia

Los objetivos básicos del análisis de supervivencia son tres:

- 1 Estimar e interpretar la función de supervivencia o la función de riesgo.
  - Curvas de Kaplan-Meier.

# Objetivos del análisis de supervivencia

Los objetivos básicos del análisis de supervivencia son tres:

- 1 Estimar e interpretar la función de supervivencia o la función de riesgo.
  - Curvas de Kaplan-Meier.
- 2 Comparar las funciones de supervivencia o de riesgo de dos o más grupos.
  - Test Log-Rank o test de Wilcoxon generalizado

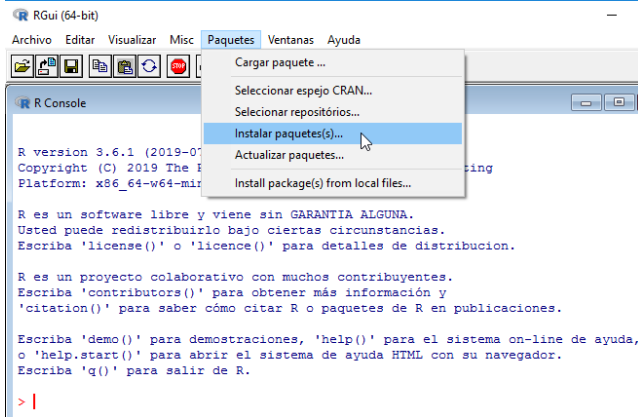
# Objetivos del análisis de supervivencia

Los objetivos básicos del análisis de supervivencia son tres:

- ① Estimar e interpretar la función de supervivencia o la función de riesgo.
  - Curvas de Kaplan-Meier.
- ② Comparar las funciones de supervivencia o de riesgo de dos o más grupos.
  - Test Log-Rank o test de Wilcoxon generalizado
- ③ Establecer la posible relación de covariables con los tiempos de supervivencia.
  - Regresión de Cox.

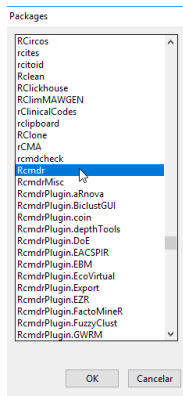
# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



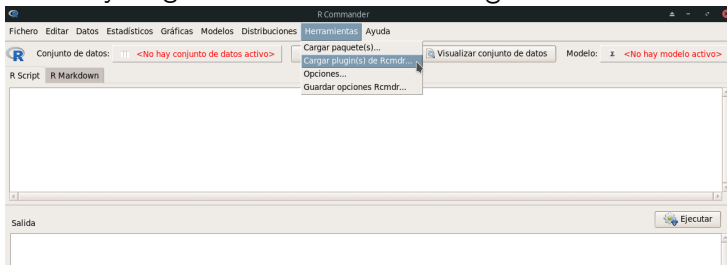
# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



# Antes de continuar...

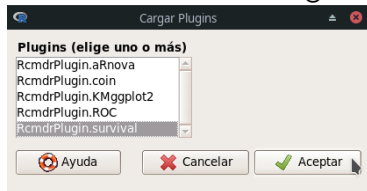
Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.





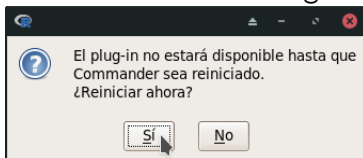
# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



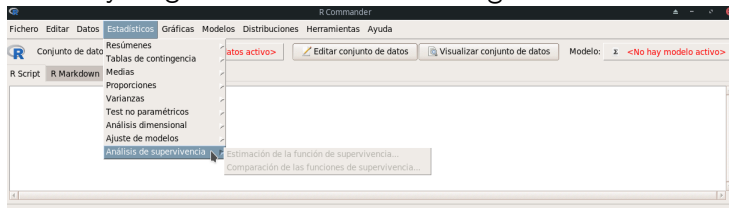
# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



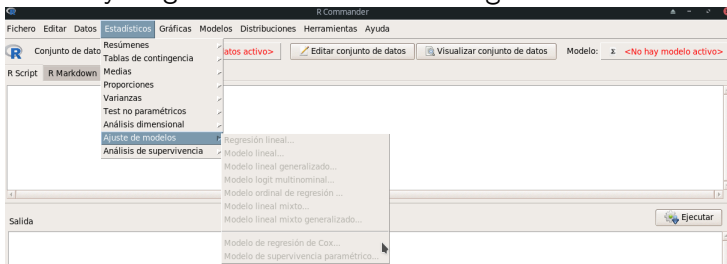
# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



# Antes de continuar...

Para poder realizar análisis de supervivencia desde R Commander es preciso instalar y cargar la extensión `RcmdrPlugin.survival`.



Estimar las funciones de supervivencia y de riesgo

# Estimación de la curva de supervivencia: método Kaplan-Meier (1/2)

Una forma de análisis de supervivencia sería fijar un horizonte  $t$  a partir del cual los datos censurados no podrían emplearse.

# Estimación de la curva de supervivencia: método Kaplan-Meier (1/2)

Una forma de análisis de supervivencia sería fijar un horizonte  $t$  a partir del cual los datos censurados no podrían emplearse.

$$\bar{\pi}_t = \frac{r_t}{n_t}$$

$$SE(\bar{\pi}_t) = \sqrt{\bar{\pi}_t(1 - \bar{\pi}_t)/n_t}$$

# Estimación de la curva de supervivencia: método Kaplan-Meier (1/2)

Una forma de análisis de supervivencia sería fijar un horizonte  $t$  a partir del cual los datos censurados no podrían emplearse.

$$\bar{\pi}_t = \frac{r_t}{n_t}$$

$$SE(\bar{\pi}_t) = \sqrt{\bar{\pi}_t(1 - \bar{\pi}_t)/n_t}$$

- Pero ¿por qué solo un horizonte temporal y no tantos como instantes en los que aparecen eventos?



# Estimación de la curva de supervivencia: método Kaplan-Meier (1/2)

Una forma de análisis de supervivencia sería fijar un horizonte  $t$  a partir del cual los datos censurados no podrían emplearse.

$$\bar{\pi}_t = \frac{r_t}{n_t}$$

$$SE(\bar{\pi}_t) = \sqrt{\bar{\pi}_t(1 - \bar{\pi}_t)/n_t}$$

- Pero ¿por qué solo un horizonte temporal y no tantos como instantes en los que aparecen eventos?

El método de Kaplan-Meier calcula la probabilidad de sobrevivir actualizando la estimación de la función de supervivencia en cada momento en que aparece un evento, sin desechar ningún dato censurado.

# Estimación de la curva de supervivencia: método Kaplan-Meier (1/2)

Una forma de análisis de supervivencia sería fijar un horizonte  $t$  a partir del cual los datos censurados no podrían emplearse.

$$\bar{\pi}_t = \frac{r_t}{n_t}$$

$$SE(\bar{\pi}_t) = \sqrt{\bar{\pi}_t(1 - \bar{\pi}_t)/n_t}$$

- Pero ¿por qué solo un horizonte temporal y no tantos como instantes en los que aparecen eventos?

El método de Kaplan-Meier calcula la probabilidad de sobrevivir actualizando la estimación de la función de supervivencia en cada momento en que aparece un evento, sin desechar ningún dato censurado.

- Con este método también es posible comparar dos o más curvas asociadas a agrupaciones de los datos.

# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

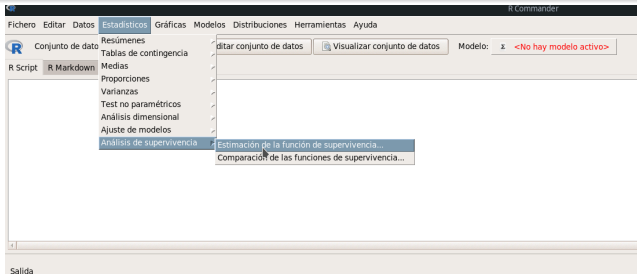
## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.

# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.



# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.

The screenshot shows a dialog box titled 'Función de supervivencia' with two tabs: 'Datos' and 'Opciones'. The 'Opciones' tab is active. It contains several sections for configuring the survival function calculation:

- Tiempo o tiempos de comienzo/final (seleccione uno o dos):** A list box with 'status' and 'time' selected.
- Indicador de evento (seleccione uno o más):** A list box with 'status' and 'time' selected.
- Estratos (seleccione cero o más):** A text box containing 'x'.
- Tipo de censura:** A group box with five radio buttons: 'Defecto' (selected), 'Derecha', 'Izquierda', 'Intervalo', and 'Intervalo tipo 2'. There is also an unchecked 'Counting process' option.
- Expresión de selección:** A text box containing '<todos los casos válidos>'.

At the bottom of the dialog are five buttons: 'Ayuda' (with a question mark icon), 'Reiniciar' (with a circular arrow icon), 'Aplicar' (with a green checkmark icon), 'Cancelar' (with a red X icon), and 'Aceptar' (with a green checkmark icon).

# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.

The screenshot shows the 'Función de supervivencia' (Survival Function) dialog box in R Studio, with the 'Opciones' (Options) tab selected. The dialog is configured for the 'leukemia' dataset. The 'Intervalos de confianza' (Confidence Intervals) section has 'Log' selected. The 'Gráfica de intervalos de confianza' (Confidence Interval Plot) section has 'Defecto' (Default) selected. The 'Nivel de confianza' (Confidence Level) is set to 0.95. The 'Método' (Method) section has 'Kaplan-Meier' selected. The 'Método de la varianza' (Variance Method) section has 'Greenwood' selected. The 'Quantiles a estimar' (Quantiles to estimate) field is set to '.25, .5, .75'. The 'Marcar tiempos de censura' (Mark censoring times) checkbox is checked. The 'Resumen' (Summary) section has 'Default' selected. At the bottom, there are five buttons: 'Ayuda' (Help), 'Reiniciar' (Reset), 'Aplicar' (Apply), 'Cancelar' (Cancel), and 'Aceptar' (Accept).

Función de supervivencia

Datos Opciones

**Intervalos de confianza**

- ☒ Log
- ☐ Log-log
- ☐ Plain
- ☐ None

**Gráfica de intervalos de confianza**

- ☒ Defecto
- ☐ Sí
- ☐ No

Nivel de confianza: 0.95

**Método**

- ☒ Kaplan-Meier
- ☐ Fleming-Harrington
- ☐ Fleming-Harrington 2

**Método de la varianza**

- ☒ Greenwood
- ☐ Tsiatis

Quantiles a estimar: .25, .5, .75

Marcar tiempos de censura ☒

**Resumen**

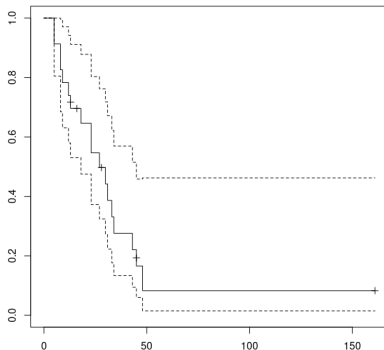
- ☒ Default
- ☐ Detailed

Ayuda Reiniciar Aplicar Cancelar Aceptar

# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.



# Estimación de la curva de supervivencia: método Kaplan-Meier (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia global.

```
> .Survfit
Call: survfit(formula = Surv(time, status) ~ 1, data = leukemia, error = "greenwood",
  conf.type = "log", conf.int = 0.95, type = "kaplan-meier")

      n  events  median 0.95LCL 0.95UCL
23      18      27      18      45

> plot(.Survfit, mark.time = TRUE)

> quantile(.Survfit, quantiles = c(0.25, 0.5, 0.75))
$quantile
25 50 75
12 27 43

$lower
25 50 75
8 18 31

$upper
25 50 75
30 48 NA
```



Comparar funciones de supervivencia y de riesgo

# Test para comparación de curvas de supervivencia (1/2)

Existen varias pruebas para contrastar si hay diferencias en las curvas de supervivencia, aunque las dos más utilizadas son:

# Test para comparación de curvas de supervivencia (1/2)

Existen varias pruebas para contrastar si hay diferencias en las curvas de supervivencia, aunque las dos más utilizadas son:

- ① Test Log-Rank o test de Mantel-Haenszel: compara las probabilidades de supervivencia en unos tiempos concretos (mediante tablas de contingencia), y combina las tablas en un solo test.

# Test para comparación de curvas de supervivencia (1/2)

Existen varias pruebas para contrastar si hay diferencias en las curvas de supervivencia, aunque las dos más utilizadas son:

- 1 Test Log-Rank o test de Mantel-Haenszel: compara las probabilidades de supervivencia en unos tiempos concretos (mediante tablas de contingencia), y combina las tablas en un solo test.
- 2 Test de Breslow o test de Wilcoxon generalizado: se trata de una generalización del test U de Mann-Whitney (o Kruskal-Wallis si ha más de dos grupos), que otorga más peso a la primera parte de la curva de supervivencia.

# Test para comparación de curvas de supervivencia (1/2)

Existen varias pruebas para contrastar si hay diferencias en las curvas de supervivencia, aunque las dos más utilizadas son:

- 1 Test Log-Rank o test de Mantel-Haenszel: compara las probabilidades de supervivencia en unos tiempos concretos (mediante tablas de contingencia), y combina las tablas en un solo test.
- 2 Test de Breslow o test de Wilcoxon generalizado: se trata de una generalización del test U de Mann-Whitney (o Kruskal-Wallis si ha más de dos grupos), que otorga más peso a la primera parte de la curva de supervivencia.

## ¿Qué test usar?

Si los distintos grupos tienen funciones de riesgo proporcionales se recomienda el test Log-Rank, y en caso contrario el test de Breslow.

## Test para comparación de curvas de supervivencia (2/2)

### El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).

# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).

The screenshot shows a dialog box titled "Función de supervivencia" with two tabs: "Datos" and "Opciones". The "Opciones" tab is active. It contains several sections for configuring the survival function calculation:

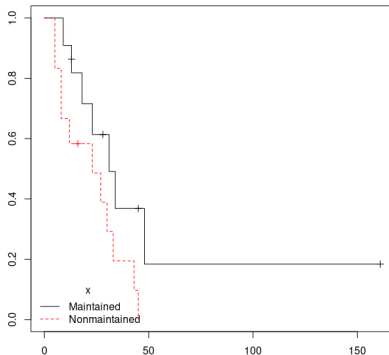
- Tiempo o tiempos de comienzo/final (seleccione uno o dos):** Two dropdown menus, both set to "time".
- Indicador de evento (seleccione uno o más):** Two dropdown menus, both set to "status".
- Estratos (seleccione cero o más):** A list box containing "x".
- Tipo de censura:** A group of radio buttons with "Defecto" selected. Other options are "Derecha", "Izquierda", "Intervalo", "Intervalo tipo 2", and "Counting process".
- Expresión de selección:** A text field containing "<todos los casos válidos>".

At the bottom of the dialog are five buttons: "Ayuda" (with a lifebuoy icon), "Reiniciar" (with a circular arrow icon), "Aplicar" (with a green checkmark icon), "Cancelar" (with a red X icon), and "Aceptar" (with a green checkmark icon).

# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).





# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).

```
x=Maintained      n events median 0.95LCL 0.95UCL
x=Nonmaintained 12      11      23      8      NA

> plot(Survfit, col = 1:2, lty = 1:2, mark.time = TRUE)

> legend("bottomleft", legend = c("Maintained", "Nonmaintained"), title = "x", col = 1:2, lty = 1:2, bty = "n")

> quantile(.Survfit, quantiles = c(0.25, 0.5, 0.75))
$quantile
      25 50 75
x=Maintained 18 31 48
x=Nonmaintained 8 23 33

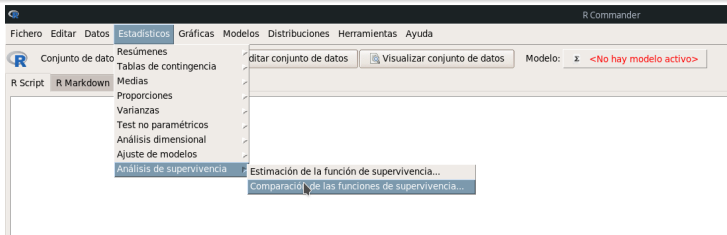
$slower
      25 50 75
x=Maintained 13 18 34
x=Nonmaintained 5 8 27

$supper
      25 50 75
x=Maintained  NA NA NA
x=Nonmaintained 30 NA NA
```

# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).



# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).

The screenshot shows a dialog box titled "Comparación de funciones de supervivencia". It contains the following fields and controls:

- Tiempo para evento (seleccione uno):** A dropdown menu with "status" and "time" options. "time" is currently selected.
- Indicador de evento (seleccione uno):** A dropdown menu with "status" and "time" options. "status" is currently selected.
- Estratos (seleccione uno o más):** A list box containing "X".
- rho:** A numeric input field with the value "0.0".
- Expresión de selección:** A text box containing "<todos los casos válidos>".
- Buttons:** "Ayuda" (Help), "Reiniciar" (Reset), "Cancelar" (Cancel), and "Aceptar" (Accept).

# Test para comparación de curvas de supervivencia (2/2)

## El problema

En la base de datos 'leukemia' (dentro del paquete 'survival') se quiere calcular la función de supervivencia según el grupo de tratamiento (Qx mantenida, Qx no mantenida).

```
> survdiff(Surv(time, status) ~ x, rho = 0, data = leukemia)
```

```
Call:
```

```
survdiff(formula = Surv(time, status) ~ x, data = leukemia, rho = 0)
```

	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
x=Maintained	11	7	10.69	1.27	3.4
x=Nonmaintained	12	11	7.31	1.86	3.4

```
Chisq= 3.4 on 1 degrees of freedom, p= 0.07
```

## Relación de covariables con tiempos de supervivencia

# Regresión de Cox (1/2)

La regresión de Cox es un modelo semi-paramétrico basado en la modelización de la función de riesgo.

# Regresión de Cox (1/2)

La regresión de Cox es un modelo semi-paramétrico basado en la modelización de la función de riesgo.

$$h(t|x) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_k x_k)$$

# Regresión de Cox (1/2)

La regresión de Cox es un modelo semi-paramétrico basado en la modelización de la función de riesgo.

$$h(t|x) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_k x_k)$$

- Es semi-paramétrica porque tiene una parte no paramétrica (función de riesgo base) y otra paramétrica (influenciada por covariables).



# Regresión de Cox (1/2)

La regresión de Cox es un modelo semi-paramétrico basado en la modelización de la función de riesgo.

$$h(t|x) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_k x_k)$$

- Es semi-paramétrica porque tiene una parte no paramétrica (función de riesgo base) y otra paramétrica (influenciada por covariables).
- Lo destacable de la regresión de Cox es que el tiempo solo aparece en el riesgo basal, de forma que las funciones de riesgo en distintos pacientes serán proporcionales.

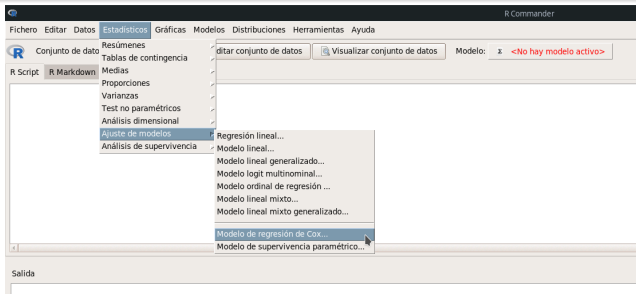
### El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.

# Regresión de Cox (2/2)

## El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.



# Regresión de Cox (2/2)

## El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.

Modelo de regresión-Cox

Datos   Modelo

**Tiempo o tiempos de comienzo/final**  
(seleccione uno o dos)

age  
bmi  
event  
p14  
toevent  
waist

**Indicador de evento**  
(seleccione uno)

age  
bmi  
event  
p14  
toevent  
waist

**Tipo de censura**

☒ Default  
☐ Right  
☐ Left  
☐ Counting process

**Estratos**  
(seleccione cero o más)

diab  
famhist  
group  
hormo  
htn  
hyperchol

**Grupos (opcional)**

diab  
famhist  
group  
hormo  
htn  
hyperchol

**Expresión de selección**  
<todos los casos válidos>

Ayuda   Reiniciar   Aplicar   Cancelar   Aceptar

# Regresión de Cox (2/2)

## El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.

Modelo de regresión-Cox

Datos | Modelo

Introduzca el nombre del modelo:

**Método para empates**

- ☒ Efron
- ☐ Breslow
- ☐ Exacto

**Errores Estándar Robustos**

- ☒ Defecto
- ☐ Sí
- ☐ No

**Variables (doble clic para enviar a la fórmula)**

- p14
- sex [factor]
- smoke [factor]
- toevent
- waist
- wth

**Fórmula del modelo**

Operadores (pulse para formular): + \* : / %in% - ^ ( )

Splines/Polinomiales: (seleccionar variable y pulsar)

B-spline spline natural polinomio ortogonal polinomio crudo

gl de los splines: 5

g. de polinomios: 2

# Regresión de Cox (2/2)

## El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.

```
> summary(modelo_cox1)
Call:
coxph(formula = Surv(toevent, event) ~ group + sex + smoke +
      with + diab + famhist + hyperchol + p14, data = predimed,
      method = "efron")

n= 6324, number of events= 252

              coef exp(coef) se(coef)      z Pr(>|z|)
group[T.MedDiet + Nuts] -0.40866  0.66454  0.15779 -2.590 0.009599 **
group[T.MedDiet + V00] -0.32355  0.72357  0.14930 -2.167 0.030228 *
sex[T.Male]             0.45564  1.57718  0.17293  2.635 0.008417 **
smoke[T.Former]         -0.01871  0.98146  0.17945 -0.104 0.916960
smoke[T.Never]          -0.45416  0.63498  0.20263 -2.241 0.025003 *
with                    1.80786  6.09741  0.99365  1.819 0.068848 .
diab[T.Yes]             0.53090  1.70046  0.13775  3.854 0.000116 ***
famhist[T.Yes]          0.13243  1.14159  0.16009  0.827 0.408116
hyperchol[T.Yes]        -0.23634  0.78951  0.13319 -1.774 0.075995 .
p14                     -0.11373  0.89250  0.03144 -3.617 0.000298 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Regresión de Cox (2/2)

## El problema

En la base de datos 'predimed' se quiere estudiar la posible asociación con la supervivencia de los grupos de intervención (3 categorías), controlando por varias covariables.

	exp(coef)	exp(-coef)	lower .95	upper .95
group[T.MedDiet + Nuts]	0.6645	1.5048	0.4878	0.9054
group[T.MedDiet + V00]	0.7236	1.3820	0.5400	0.9695
sex[T.Male]	1.5772	0.6340	1.1238	2.2135
smoke[T.Former]	0.9815	1.0189	0.6904	1.3952
smoke[T.Never]	0.6350	1.5748	0.4269	0.9446
wth	6.0974	0.1640	0.8697	42.7508
diab[T.Yes]	1.7005	0.5881	1.2981	2.2275
famhist[T.Yes]	1.1416	0.8760	0.8342	1.5623
hyperchol[T.Yes]	0.7895	1.2666	0.6081	1.0250
p14	0.8925	1.1204	0.8392	0.9492

Concordance= 0.673 (se = 0.017 )  
Likelihood ratio test= 91.48 on 10 df, p=3e-15  
Wald test = 89.83 on 10 df, p=6e-15  
Score (logrank) test = 92.76 on 10 df, p=2e-15

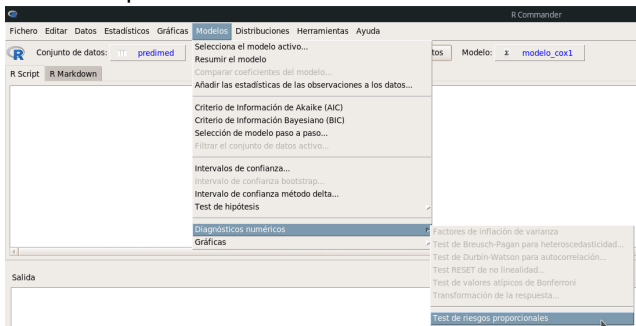
## Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



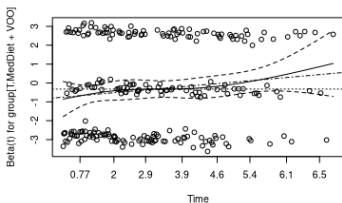
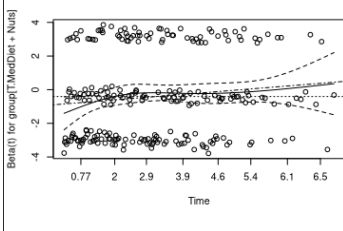
# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



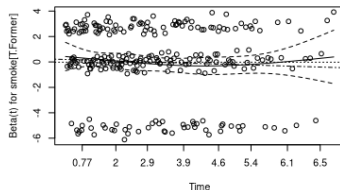
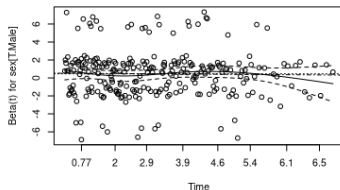
# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



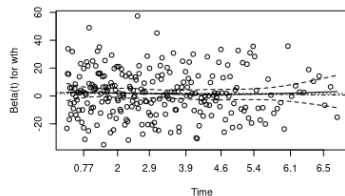
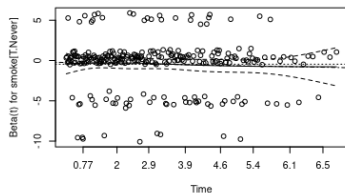
# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



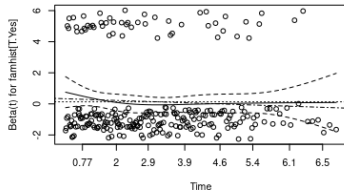
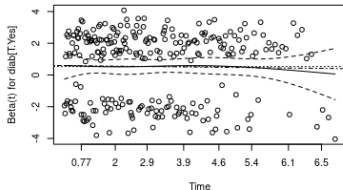
# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



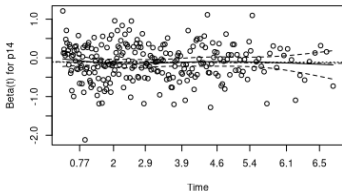
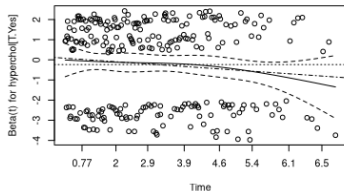
# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.



# Riesgos proporcionales: test y gráfico

No siempre resulta verosímil asumir riesgos proporcionales, de forma que hay que evaluar este punto del modelo.

```
> .CoxZPH
```

	rho	chisq	p
group[T.MedDiet + Nuts]	0.1230	3.8390	0.0501
group[T.MedDiet + V00]	0.1275	4.0782	0.0434
sex[T.Male]	-0.0182	0.0754	0.7836
smoke[T.Former]	-0.0526	0.7014	0.4023
smoke[T.Never]	-0.0460	0.5381	0.4632
wth	-0.0245	0.1613	0.6880
diab[T.Yes]	-0.0203	0.1103	0.7398
famhist[T.Yes]	-0.0543	0.7510	0.3861
hyperchol[T.Yes]	-0.1044	2.8060	0.0939
p14	-0.0198	0.0911	0.7628
GLOBAL	NA	10.3013	0.4145

## ¿Qué hacer si no hay riesgos proporcionales?

La solución más completa será introducir dependencia temporal en las covariables, aunque lo más habitual (y sencillo) es considerar a esa covariable como un estrato de los datos.



# ¿Qué hacer si no hay riesgos proporcionales?

La solución más completa será introducir dependencia temporal en las covariables, aunque lo más habitual (y sencillo) es considerar a esa covariable como un estrato de los datos.

Modelo de regresión-Cox

Datos    Modelo

**Tiempo o tiempos de comienzo/final  
(seleccione uno o dos)**

- age
- bmi
- event
- p14
- toevent
- waist

**Indicador de evento  
(seleccione uno)**

- age
- bmi
- event
- p14
- toevent
- waist

**Tipo de censura**

- ☒ Default
- ☐ Right
- ☐ Left
- ☐ Counting process

**Estratos  
(seleccione cero o más)**

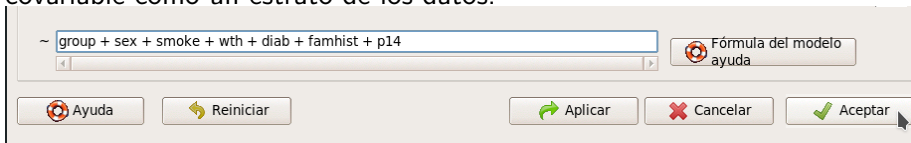
- diab
- famhist
- group
- hormo
- htn
- hyperchol

**Grupos  
(opcional)**

- diab
- famhist
- group
- hormo
- htn
- hyperchol

## ¿Qué hacer si no hay riesgos proporcionales?

La solución más completa será introducir dependencia temporal en las covariables, aunque lo más habitual (y sencillo) es considerar a esa covariable como un estrato de los datos.



The screenshot shows a software interface for a statistical model. At the top, there is a text input field containing the formula: `~ group + sex + smoke + wth + diab + famhist + p14`. To the right of the input field is a button labeled "Fórmula del modelo ayuda" with a lifebuoy icon. Below the input field, there are four buttons: "Ayuda" (with a lifebuoy icon), "Reiniciar" (with a circular arrow icon), "Aplicar" (with a green curved arrow icon), and "Cancelar" (with a red X icon). To the right of these is a button labeled "Aceptar" with a green checkmark icon. A mouse cursor is pointing at the "Aceptar" button.

# ¿Qué hacer si no hay riesgos proporcionales?

La solución más completa será introducir dependencia temporal en las covariables, aunque lo más habitual (y sencillo) es considerar a esa covariable como un estrato de los datos.

```
> CoxModel.4 <- coxph(Surv(toevent, event) ~ group + sex + smoke + wth + diab + famhist + p14 + strata(hyperchol), method = "efron", data = predimed)
> summary(CoxModel.4)
Call:
coxph(formula = Surv(toevent, event) ~ group + sex + smoke +
      wth + diab + famhist + p14 + strata(hyperchol), data = predimed,
      method = "efron")

n= 6324, number of events= 252

              coef exp(coef) se(coef)      z Pr(>|z|)
group[T.MedDiet + Nuts] -0.41067   0.66321  0.15779 -2.603 0.009251 **
group[T.MedDiet + V00] -0.32539   0.72224  0.14932 -2.179 0.029319 *
sex[T.Male]             0.45526   1.57659  0.17311  2.630 0.008540 **
smoke[T.Former]         -0.02341   0.97686  0.17946 -0.130 0.896200
smoke[T.Never]          -0.45732   0.63298  0.20270 -2.256 0.024059 *
wth                     1.80467   6.07796  0.99287  1.818 0.069120 .
diab[T.Yes]             0.53360   1.70506  0.13780  3.872 0.000108 ***
famhist[T.Yes]          0.13315   1.14242  0.16004  0.832 0.405402
p14                     -0.11350   0.89270  0.03149 -3.604 0.000313 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## ¿Qué hacer si no hay riesgos proporcionales?

La solución más completa será introducir dependencia temporal en las covariables, aunque lo más habitual (y sencillo) es considerar a esa covariable como un estrato de los datos.

```
> .CoxZPH <- cox.zph(CoxModel.4)
```

```
> .CoxZPH
```

	rho	chisq	p
group[T.MedDiet + Nuts]	0.1222	3.7927	0.0515
group[T.MedDiet + V00]	0.1253	3.9346	0.0473
sex[T.Male]	-0.0190	0.0824	0.7740
smoke[T.Former]	-0.0548	0.7623	0.3826
smoke[T.Never]	-0.0470	0.5630	0.4531
wth	-0.0248	0.1651	0.6845
diab[T.Yes]	-0.0187	0.0935	0.7598
famhist[T.Yes]	-0.0544	0.7535	0.3854
p14	-0.0188	0.0821	0.7745
GLOBAL	NA	7.5240	0.5827