

Introducción al Análisis de Supervivencia

Introducción

En este tema abordamos el estudio del *tiempo transcurrido entre dos eventos de interés* y de las variables que pueden influir en el mismo. Por ejemplo, podemos querer estimar el tiempo que transcurre desde que aplicamos un tratamiento hasta que se produce la curación total de un paciente, el tiempo que transcurre desde que a una persona se le diagnostica una determinada enfermedad hasta que fallece a causa de la misma, o el tiempo que transcurre desde que realizamos un trasplante de un órgano hasta que el órgano deja de funcionar,...Además, nos puede interesar conocer cuál es el efecto que sobre el tiempo que tarda en curar un paciente tiene la edad del mismo, o si el tiempo desde el diagnóstico hasta la muerte depende de factores como la edad al diagnóstico, el sexo,...En principio, tratamos de modelizar una variable cuantitativa, pues estamos hablando de “tiempos”, y a continuación explicamos cuál es el motivo por el cuál no se puede abordar su modelización desde la perspectiva de la Regresión Lineal. En primer lugar, la variable *tiempo* no acostumbra a seguir una distribución Normal, y en segundo lugar, en algunas ocasiones no tenemos la posibilidad de observar el evento de interés. Por ejemplo, en un estudio sobre el tiempo hasta la muerte de pacientes diagnosticados de una determinada enfermedad, al finalizar el estudio algunos pacientes habrán fallecido por la causa en estudio, otros seguirán vivos e incluso a otros podríamos haberles perdido el seguimiento o podrían haber fallecido por una causa diferente a la que es objeto de estudio (por ejemplo en un *accidente de tráfico*). Sin embargo, sabemos el tiempo transcurrido hasta la muerte en aquellos pacientes que han fallecido por la causa de interés y del resto sabemos que al menos han sobrevivido a la enfermedad hasta la última

vez que los observamos con vida. Si únicamente se incluyen en el análisis aquellos pacientes que han fallecido por la causa podríamos obtener resultados sesgados y además estaríamos omitiendo información válida sobre el resto.

Así, en algunas observaciones se habrá producido el evento de interés en el momento de finalización del estudio y en otras, aunque no dispondremos del tiempo transcurrido hasta el evento de interés, tendremos información parcial sobre cuál es el último momento hasta el cuál tenemos conocimiento de que no se había producido. A estas últimas observaciones se les llama **datos censurados**. Pueden ser diferentes las causas de censura, pero siempre responden a alguno de los siguientes criterios:

- 1.- existe una pérdida de seguimiento (por ejemplo, el paciente desaparece o decide no seguir participando en el estudio).
- 2.- el estudio termina antes de que se produzca el evento
- 3.- se produce otro evento que impide que ocurra el evento de interés (por ejemplo, estamos observando el tiempo desde el diagnóstico de una enfermedad hasta la muerte por esa enfermedad y el paciente fallece en un accidente de tráfico).

Para cada individuo tendremos, por tanto, la siguiente información:

- **Fecha inicio:** la fecha de comienzo del seguimiento (fecha de diagnóstico de la enfermedad, fecha del trasplante,...)
- **Fecha fin:** la fecha en la que le ocurre el evento, o para los individuos a los que no les ocurre el evento la última fecha conocida en la que al paciente no le había ocurrido el evento.
- A partir de estas dos fechas calcularemos el **tiempo T** a modelizar
- **Evento:** notificación sobre si la fecha anterior se refiere a la ocurrencia del evento o última fecha en que sigue sin ocurrir (por ejemplo muerto/vivo, curación/no curación, ...). A este dato se le suele llamar **censura**, ya que indica qué datos están censurados y cuáles no (los no censurados son aquellos a los que les ocurre el evento). Se suele representar con el valor **1** a los individuos a los que les ha ocurrido el evento, y con el valor **0** a los que no.

- Otras **covariables** sobre las que queremos valorar la influencia sobre el tiempo de interés (Sexo, Edad,)

Ejemplo:

Supongamos que estamos estudiando el tiempo que transcurre desde que administramos un antibiótico por infección urinaria hasta que se produce una recaída o reinfección. Supongamos además que nuestro estudio se lleva a cabo durante 12 meses y disponemos de la información de 5 pacientes que se resume a continuación:

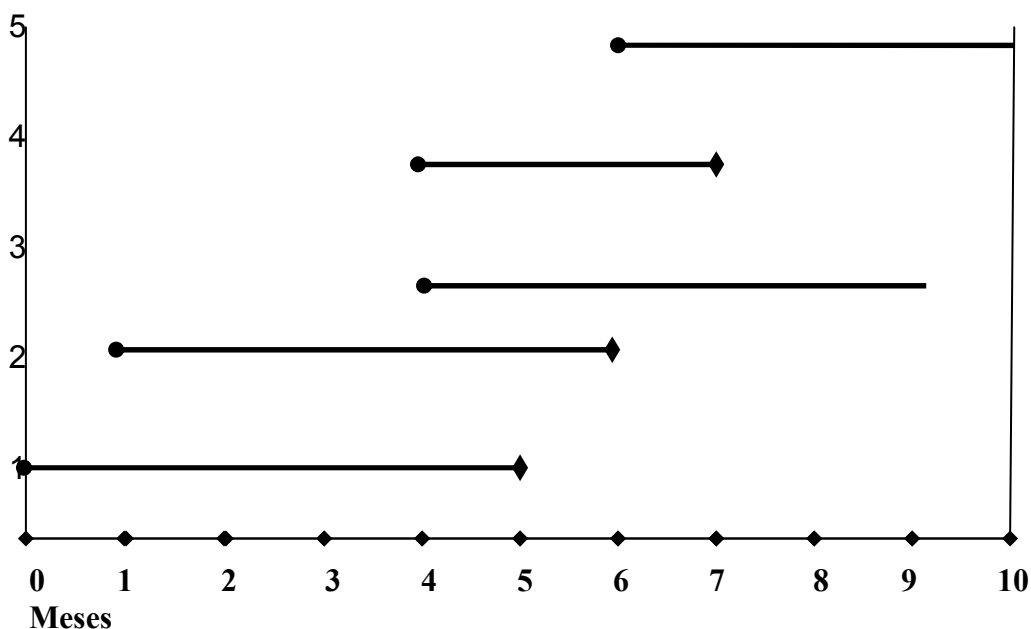


Figura 1.- Ejemplo de seguimiento de 5 individuos

Podemos observar que el individuo es tratada al inicio del estudio y en él se produce una recaída a los 5 meses. El individuo 2 es tratado en el mes 1 de la infección urinaria y se produce en él una recaída en el mes 6, por lo que también transcurren 5 meses hasta la recaída. El individuo 3 es tratado en el mes 4 y en el mes 9 dejamos de tener información sobre él (se le pierde el seguimiento), eso sí, sabemos que durante los 5 meses no tuvo una recaída. El individuo 4 también es tratado de la infección en el mes 4 y se produce en él una recaída en el mes 7, por lo que el tiempo hasta la recaída en este caso es de 3 meses. Mientras que el individuo 5 es tratado de la infección en el mes 6 y

en el mes 10 finaliza el estudio sin que en este sujeto se hubiera producido recaída, por lo que sabemos que el tiempo hasta la recaída es superior a los 4 meses. Los datos de los individuos 3 y 5 son datos **censurados**, ya que en ellos no conocemos el tiempo hasta el evento de interés, aunque sabemos que este tipo será el menos cierto número de meses sobre el que tenemos conocimiento.

A la hora de exponer este tipo de análisis se suele considerar por simplificar la explicación que el evento de interés es “la muerte” y por tanto el tiempo hasta la ocurrencia del evento se le suele llamar tiempo de supervivencia (aunque como hemos comentado este tipo de análisis se aplica también para medir *tiempo hasta curación, tiempo hasta el rechazo de un órgano,.....*). El **tiempo de supervivencia** se define entonces como el transcurrido desde la fecha de comienzo del seguimiento hasta la fecha del último contacto con el paciente, bien por haber fallecido bien por haber sido censurado.

Para planificar correctamente un estudio de supervivencia, deberemos definir apropiadamente el inicio del seguimiento, la unidad de tiempo que vamos a utilizar (semana, mes, año,...) y el evento de interés que queremos analizar.

Existen diferentes herramientas para el análisis de datos de supervivencia. Entre ellas hemos seleccionado para este curso dos herramientas que se enmarcan dentro de las llamadas pruebas no paramétricas, la metodología de Kaplan-Meier y la Regresión de Cox.

La metodología estadística que analiza el **tiempo de supervivencia** tiene objetivo obtener una función dependiente del tiempo (t), cuyos valores proporcionen la probabilidad de sobrevivir por lo menos hasta el instante t . A esta función se le llama **curva de supervivencia**.

Función de supervivencia

La **función de supervivencia** proporciona la probabilidad de que un paciente sobreviva un determinado tiempo.

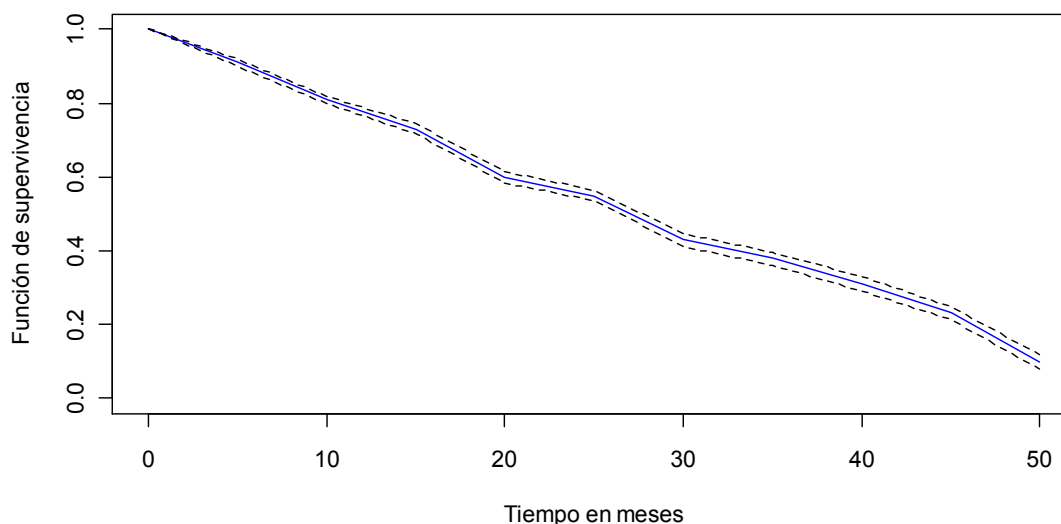


Figura 2.- Función de supervivencia con su correspondiente intervalo de confianza al 95%.

En esta figura se representa la función de supervivencia y podemos observar como varía, a lo largo del tiempo, la proporción de individuos que están libres del evento de interés. Por ejemplo, en el mes 20, alrededor del 60% de los individuos todavía no habían presentado el evento de interés. En línea discontinua se representa el intervalo de confianza al 95% para la supervivencia en cada momento. El hecho de que este intervalo sea tan estrecho indica que se dispone de buena información.

La función de supervivencia siempre mostrará un descenso rápido al inicio por el hecho de haber más individuos expuestos al riesgo que a medida que pasa el tiempo. Por este motivo no son comparables los descensos que se observan al inicio de la función de supervivencia (obtenidos de una población grande) con los obtenidos al final (cuando la población en riesgo ha disminuido considerablemente). Para obtener una función comparable se calcula la **función de riesgo**.

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. Las funciones de riesgo pueden presentar diferentes formas (no necesariamente siempre descendiente como la función de supervivencia) que responden a diferentes situaciones.

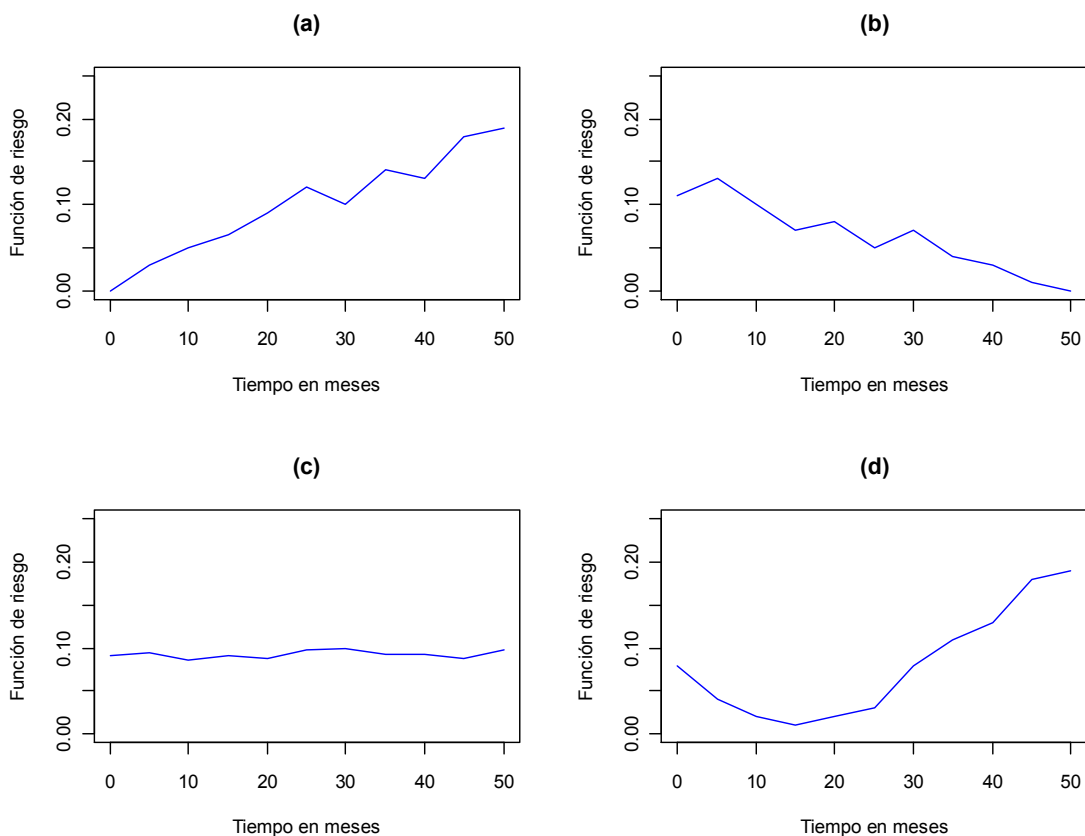


Figura 3.- Función de riesgo en 4 situaciones diferentes.

En la figura 3 se observan las funciones de riesgo que describen cuatro situaciones diferentes. La situación (a) podría responder al análisis del tiempo de vida de personas con leucemia que no responden al tratamiento, y en los que la función de riesgo es creciente. La situación (b) podría responder a la función de riesgo tras una intervención quirúrgica, en la que la función es decreciente. La situación (c) responde a situaciones de riesgo constante, como por ejemplo el riesgo de hemorragias digestivas tras la ingesta de antiinflamatorios (de los que se afirma que hay el mismo riesgo sea la primera o la enésima vez que se toman). Por último, la situación (d) describe por

ejemplo el tiempo de vida de un ser vivo, con mayor riesgo al principio y al final de la vida (tiene lo que se denomina “forma de bañera”).

El método de Kaplan-Meier

El método de Kaplan-Meier es un método que permite estimar la curva de supervivencia. Este método, enmarcado como no paramétrico, requiere muy pocas restricciones para su aplicación, y lo único que supone es que los sujetos censurados se habrían comportado del mismo modo que los seguidos hasta que se produjo el evento.

A simple vista, parece razonable estimar la proporción de pacientes que sobreviven a un determinado tiempo t como el número de individuos cuyo tiempo de observación ha sido superior o igual a t , es decir, los que “sobreviven” sobre todos los pacientes de los que disponemos información hasta por lo menos ese instante.

Para cada periodo de tiempo se calcula la probabilidad de sobrevivir, y la función de Kaplan-Meier es la “probabilidad de supervivencia individual acumulada a lo largo del tiempo”. Este método “actualiza” la estimación de la función de supervivencia en cada momento en que aparece un evento.

Mediante este método también es posible la comparación de dos o más curvas de supervivencia correspondientes a dos o más grupos (hombres/mujeres, tratamiento1/tratamiento2/tratamiento 3, ...) definidos por una variable categórica.

Regresión de Cox. Modelo de riesgos proporcionales

La regresión de Cox modeliza la función de riesgo, y es una generalización del análisis de Kaplan-Meier. Mientras que con este último método podemos realizar comparaciones entre grupos diferentes definidos por un factor simple, la regresión de Cox ofrece la posibilidad de introducir en el modelo un conjunto de covariables, categóricas o continuas. Así, la regresión de Cox, consiste en obtener una función lineal de las variables independientes que permita estimar, en función del tiempo, la probabilidad de que ocurra el evento de interés. Por ejemplo, podemos querer estimar la probabilidad de sobrevivir hasta cierto

instante t en función del sexo (Hombre/Mujer), el tratamiento recibido (1,2 y 3), edad del paciente,...

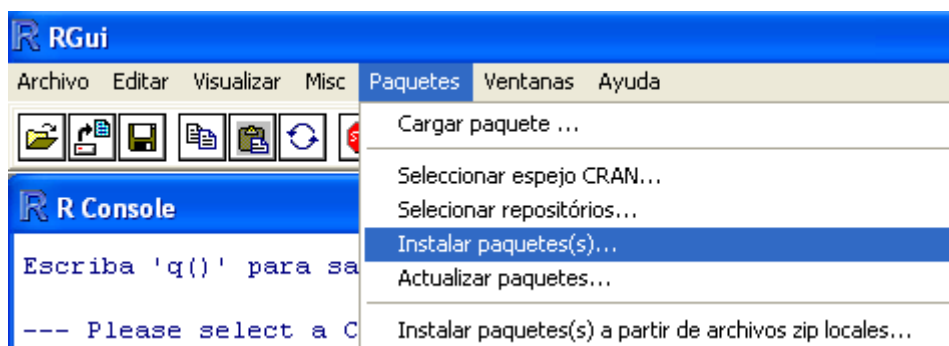
La regresión de Cox tiene como premisa que los riesgos son proporcionales (para los grupos definidos por las covariables) y el análisis con este modelo consiste en estimar unos coeficientes, como si de Regresión lineal o Regresión Logística se tratara que indicarán la influencia de las covariables en el tiempo de supervivencia.

Como las modelizaciones citadas, debe tenerse especial cuidado con la colinealidad de las variables explicativas y con el cumplimiento de las premisas necesarias para poder aplicar la modelización (aunque en este tema no profundizaremos en esta práctica, pues el objetivo es “introducir” el análisis de datos de supervivencia)

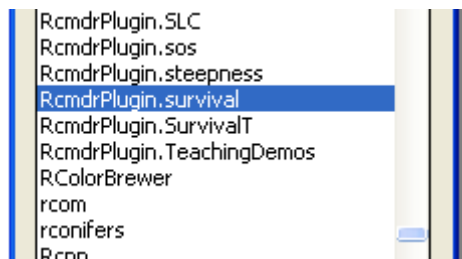
Antes de poder trabajar sobre un ejemplo, es necesario adaptar el R-Commander a la nueva funcionalidad que necesitamos.

Actualización del R-Commander para disponer de la funcionalidad de Análisis de Supervivencia

Desde la ventana de R seleccionamos la opción: **Paquetes>Instalar paquetes(s)...**

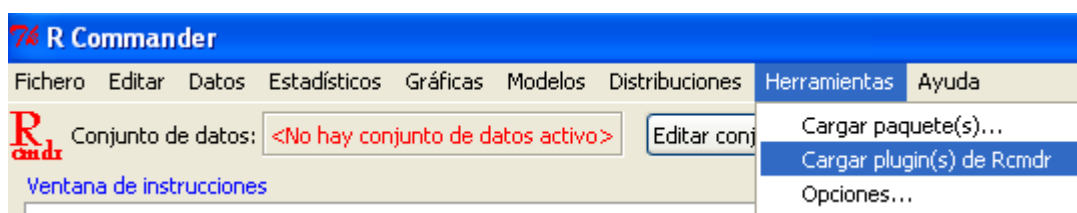


Seleccionamos el paquete “RcmdrPlugin.survival”

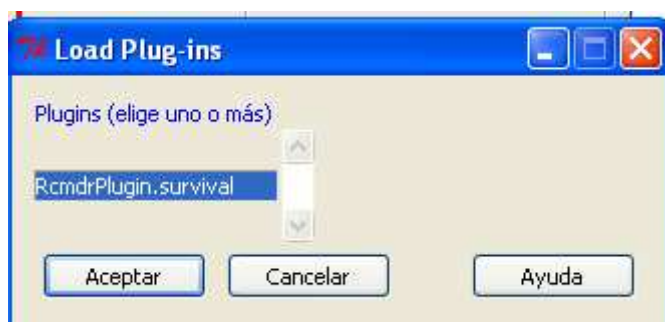


A continuación cargamos el R-Commander (tecleando como siempre en la ventana de R el comando: `library(Rcmdr)`).

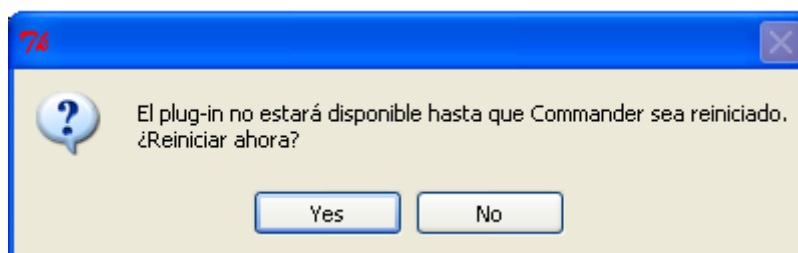
Una vez dentro de la ventana de R-Commander seleccionamos la opción Herramientas>Cargar plugin(s) de Rcmdr:



Nos aparecerá la siguiente ventana:

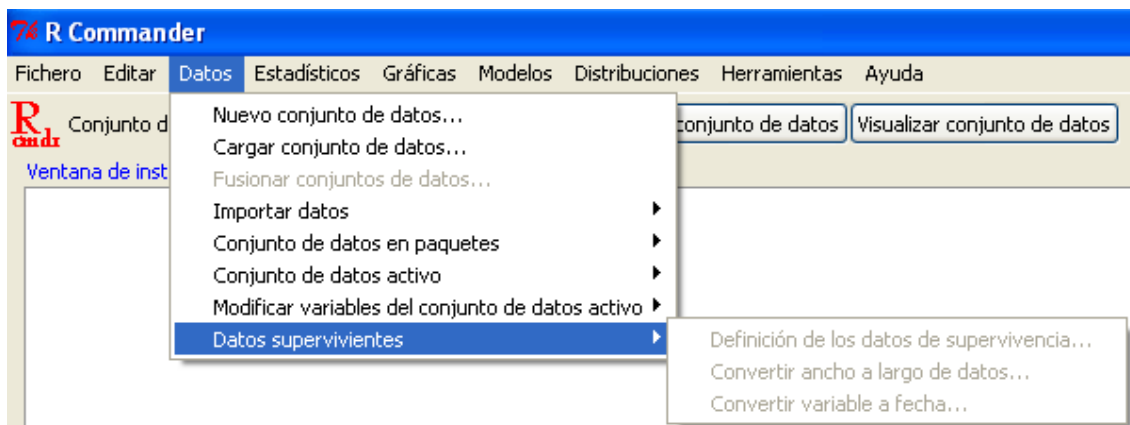


y pulsaremos “Aceptar”. A continuación nos preguntará:

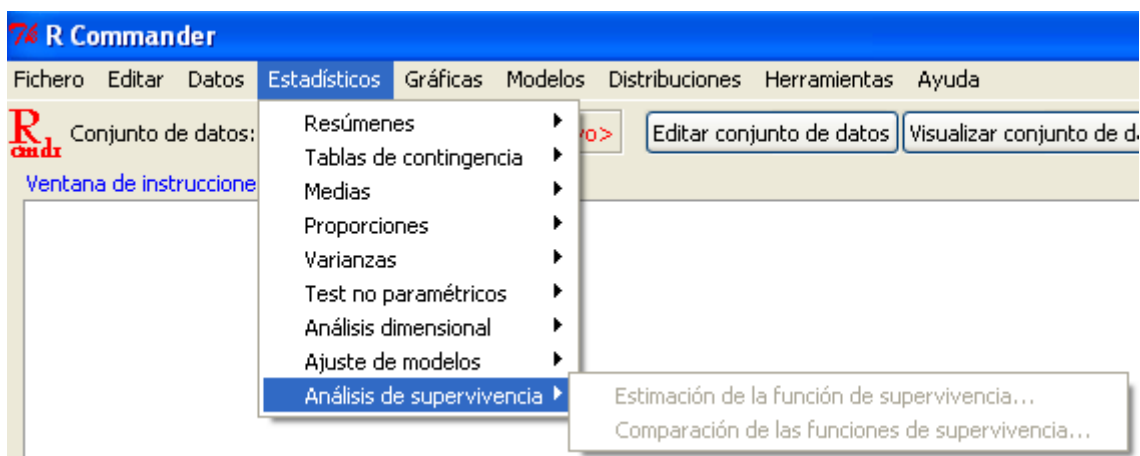


a lo que responderemos “Yes”. A partir de ese momento R-Commander se cerrará y se volverá a abrir, apareciendo en él nuevas opciones de menú. Esta última operación la tendremos que realizar cada vez que queramos disponer en R-Commander de la funcionalidad del Análisis de Supervivencia. Concretamente podremos observar que han aparecido las siguientes opciones nuevas (entre otras) en los diferentes menús del programa:

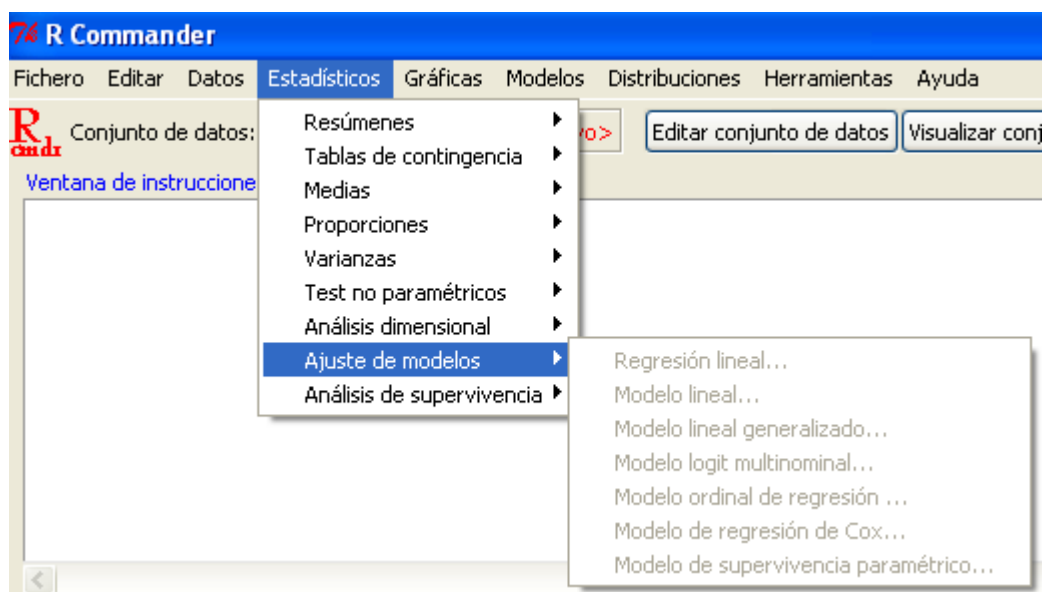
- 1.- Datos > Datos supervivientes...



2.- Estadísticos > Análisis de supervivencia...



3.- Estadísticos > Ajuste de modelos > Modelo de regresión de Cox... y Modelo de supervivencia paramétrico...



Ejemplo 1: Estimación de la curva de supervivencia (Kaplan-Meier) y comparación de dos curvas con R-Commander

Banco de datos: **Leucemia.rda**

42 niños con leucemia son divididos aleatoriamente en 2 grupos diferentes, el grupo control y el grupo que añade una nueva componente al tratamiento estándar (consideraremos que se trata de un “nuevo tratamiento”)

tiempo: tiempo hasta la recaída

censura: 1=ocurrido evento, 0=dato censurado

trata: 0 (tratamiento control) y 1 (tratamiento nuevo)

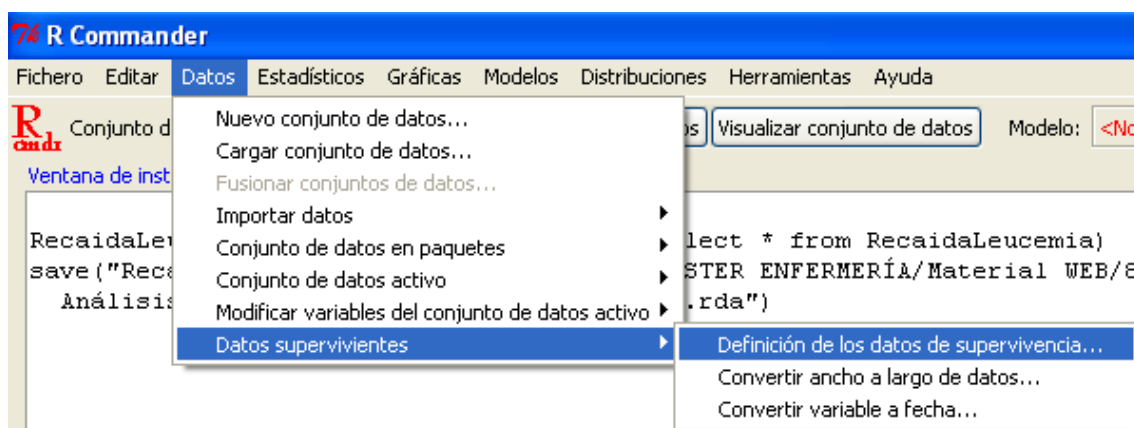
estadio: estadio en la que está la enfermedad.

▣ **Paso 1: Cargar banco de datos: Leucemia.rda** (tal y como se cargan habitualmente)

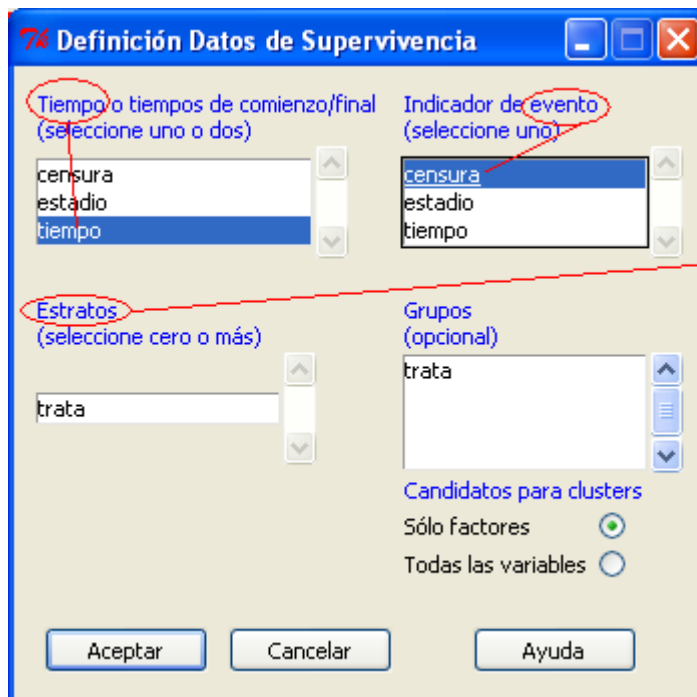
▣ **Paso 2: Indicar al programa cuál es la variable que guarda el tiempo y cuál es la que guarda la censura.**

En primer lugar vamos a indicarle al programa las dos variables que recogen de forma conjunta el evento de interés: **tiempo** y **censura**. Así, cuando queramos realizar análisis sobre estos datos por defecto el programa asignará las dos variables mencionadas de forma automática.

Vamos a la opción: Datos> Datos supervivientes> Definición de los datos de supervivencia



Y a continuación indicamos las variables tiempo y censura en sus correspondientes lugares:

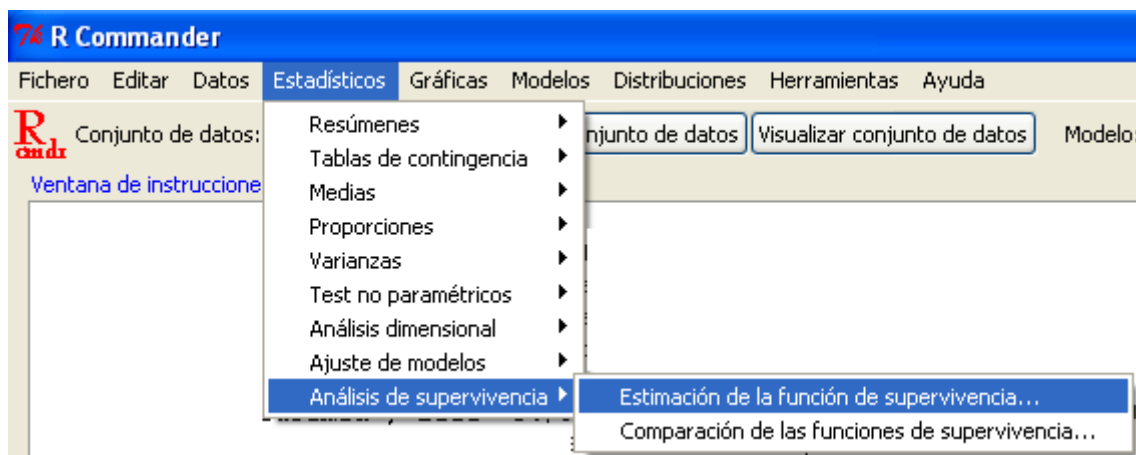


Marcaríamos trata si quisiéramos comparar las curvas de supervivencia de los que siguen ambos tratamientos. (De momento no)

Pulsamos **Aceptar** y ya tendríamos los datos definidos para los análisis siguientes.

▣ Paso 3: Estimación de la curva de supervivencia por el método de Kaplan-Meier

Vamos a la opción de menú: Estadísticos > Análisis de supervivencia > Estimación de la función de supervivencia



Nos aparecerá la siguiente ventana:

Función de supervivencia

Tiempo o tiempos de comienzo/final (seleccione uno o dos): censura, estado, tiempo

Indicador de evento (seleccione uno): censura, estado, tiempo

Estratos (seleccione cero o más): trata

Intervalos de confianza: Log (seleccionado), Log-log, Plain, None

Gráfica de intervalos de confianza: Defecto, Sí (seleccionado), No

Nivel de confianza: .95

Marcar tiempos de censura:

Método: Kaplan-Meier (seleccionado), Fleming-Harrington, Fleming-Harrington 2

Método de la varianza: Greenwood (seleccionado), Tsiatis

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

Expresión de selección: <todos los casos válidos

Botones: Aceptar, Cancelar, Ayuda

Las variables ya aparecen seleccionadas por defecto porque hemos definido la estructura de datos en el paso 2.

Si queremos que incluya en la representación gráfica los intervalos de confianza para la función de supervivencia marcamos esta opción.

Aunque existen otros métodos, el de Kaplan-Meier es el más extendido en uso.

En la que tras seleccionar los criterios deseados pulsaremos Aceptar.

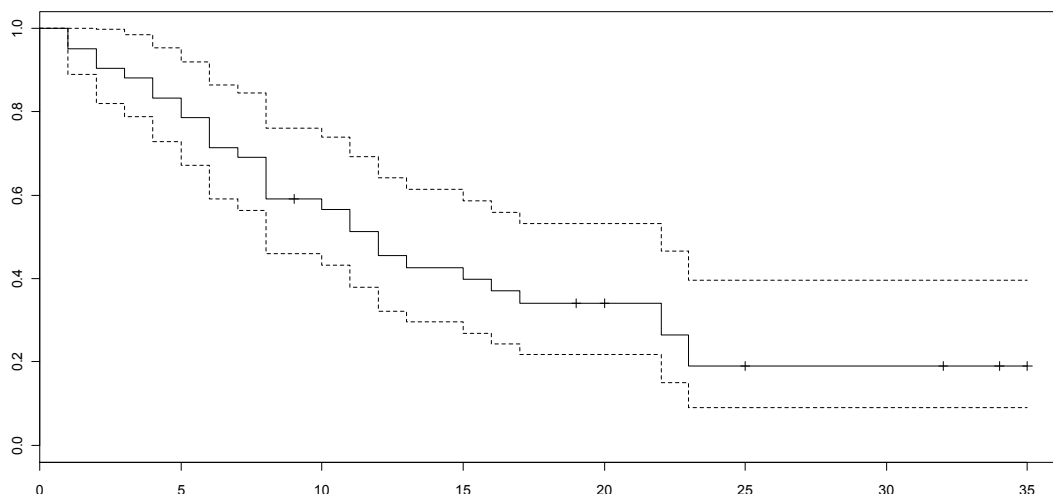
El programa devuelve, entre otras, el resultado:

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
1	42	2	0.952	0.0329	0.890	1.000		
2	40	2	0.905	0.0453	0.820	0.998		
3	38	1	0.881	0.0500	0.788	0.985		
4	37	2	0.833	0.0575	0.728	0.954		
5	35	2	0.786	0.0633	0.671	0.920		
6	33	3	0.714	0.0697	0.590	0.865		
7	29	1	0.690	0.0715	0.563	0.845		
8	28	4	0.591	0.0764	0.459	0.762		
10	23	1	0.565	0.0773	0.433	0.739		
11	21	2	0.512	0.0788	0.378	0.692		
12	18	2	0.455	0.0796	0.323	0.641		
13	16	1	0.426	0.0795	0.296	0.615		
15	15	1	0.398	0.0791	0.269	0.588		
16	14	1	0.369	0.0784	0.244	0.560		
17	13	1	0.341	0.0774	0.219	0.532		
22	9	2	0.265	0.0765	0.151	0.467		
23	7	2	0.189	0.0710	0.091	0.395		

En la primera línea podemos observar que en el primer mes, tenemos 42 individuos en seguimiento, de los cuales solo 2 tienen una recaída, mientras que 40 no la tienen. La probabilidad por tanto de “sobrevivir a la recaída” un mes es $40/42=0.952$ (esta probabilidad se acompaña del intervalo de confianza

correspondiente). En las siguientes líneas el cálculo es similar, pero también se tienen en cuenta las pérdidas de seguimiento (es decir, las censuras).

En la ventana de gráficos de R, el programa también proporciona la siguiente gráfica:



En esta figura podemos apreciar la curva de supervivencia calculada. La probabilidad de sobrevivir hasta el mes 20 está alrededor (aunque algo menor) del 40%.

En la ventana de R-Commander también aparecen los siguientes resultados:

	Survival Probability		
Estimate	0.25	0.5	0.75
lower 0.95 CL	16	8	4
quantile	23	12	6
upper 0.95 CL	NA	22	10

Probabilidades (points to 0.25, 0.5, 0.75)
Meses (points to 23, 12, 6)

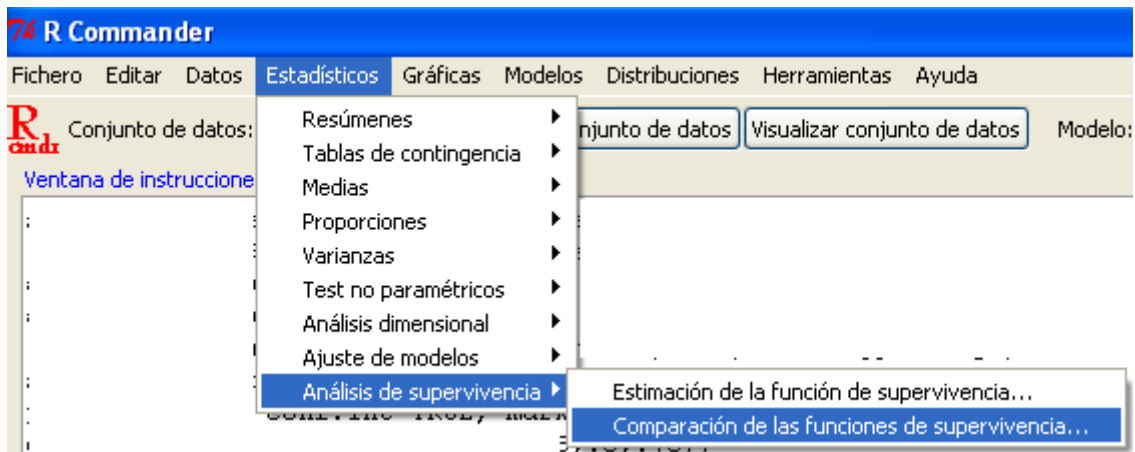
La interpretación de estos resultados es la siguiente:

- El percentil 75 del tiempo de supervivencia está alrededor de 4, con un intervalo de confianza del 95% entre 4 y 10. Esto significa que el 75% de los enfermos no tiene recaída (“sobrevive a la recaída”) al menos 6 meses (con un 95% de confianza entre 4 y 10)
- El percentil 50 del tiempo de supervivencia está alrededor de 12 (IC95%=(8,22)), por tanto, el 50% de los enfermos no tiene recaída al menos en 12 meses (con un 95% de confianza entre 8 y 22).
- Por último el 25% de los pacientes no tiene recaída hasta los 23 meses aprox. (con un 95% de confianza entre 16 y más de 23 [infinito])

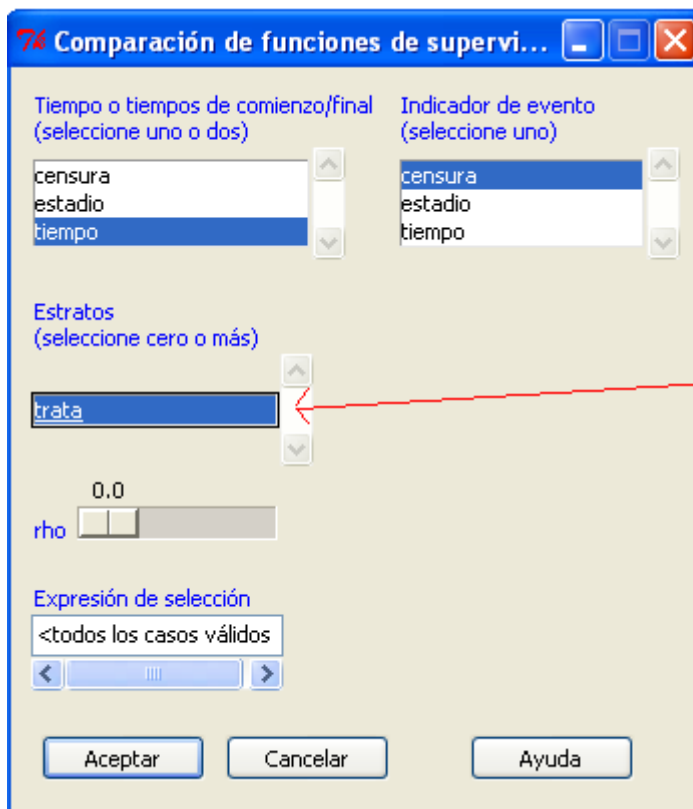
□ Paso 4: Comparación de curvas de supervivencia

Hasta ahora, hemos tratado a los 42 pacientes como si de un grupo de una misma población se tratara. La realidad es que de los 42 pacientes algunos forman el grupo control (tratamiento estándar) y otros forman un grupo a los que se les ha aplicado el tratamiento nuevo. Podría ocurrir, que el tiempo hasta la recaída de estos pacientes dependiera o tuviera relación con el tratamiento seguido.

Para comprobar esta hipótesis podemos realizar la comparación de curvas de supervivencia de ambos tipos de pacientes, los que toman el tratamiento control y los que toman el tratamiento nuevo. Para ello acudimos a la opción de menú: Estadísticos> Análisis de supervivencia> Comparación d las funciones de supervivencia:



Aparecerá la ventana:



Seleccionamos "trata" para indicar que queremos comparar las curvas de supervivencia de los estratos o grupos que define esta variable categórica.

Tras pulsar el botón Aceptar el R-Commander nos devolverá el resultado:

```
> survdiff(Surv(tiempo,censura) ~ trata, rho=0, data=RecaidaLeucemia)
Call:
survdiff(formula = Surv(tiempo, censura) ~ trata, data = RecaidaLeucemia,
         rho = 0)

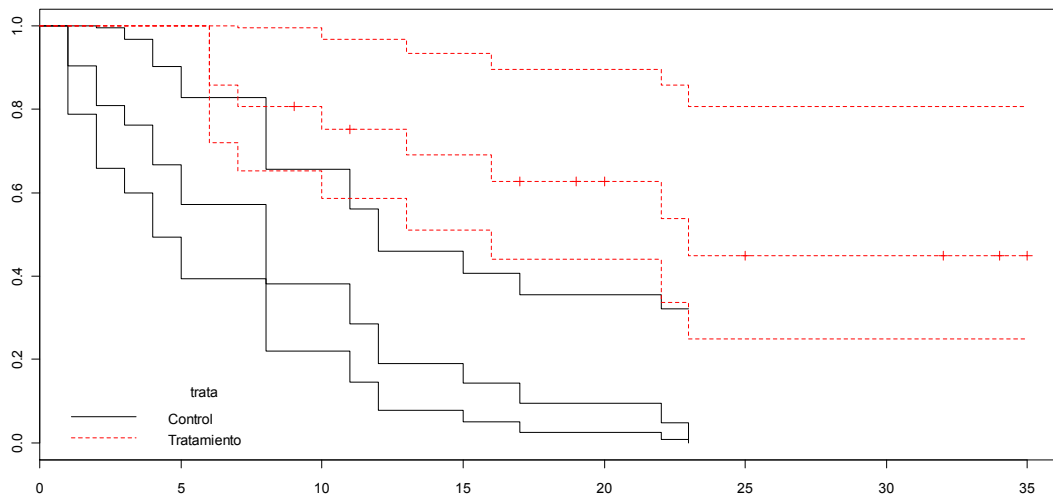
      N Observed Expected (O-E)^2/E (O-E)^2/V
trata=Control    21      21   10.7    9.77   16.8
trata=Tratamiento 21       9   19.3    5.46   16.8

Chisq= 16.8 on 1 degrees of freedom, p= 4.17e-05
```

p-valor resultante de la comparación de las dos curvas de supervivencia (donde la hipótesis nula es que ambas curvas son iguales)

Como el p-valor obtenido es 0.0000417, y es mucho menor que el nivel de significatividad habitual (0.05), rechazamos la hipótesis nula que indica igualdad de curvas de supervivencia, y por tanto podemos afirmar que la supervivencia (a la recaída) de ambos tratamientos no tiene el mismo comportamiento. Para ilustrar a nivel gráfico cómo se está comportando cada grupo y para ello, volvemos a la opción Estadísticos> Análisis de supervivencia> Estimación de la función de supervivencia, pero en esta ocasión en la opción Estratos seleccionamos la variable "trata". El programa nos devuelve la estimación de la curva de supervivencia (numérica y gráfica)

pero separadamente para ambos grupos. A continuación se muestra la salida gráfica:



Podemos apreciar que la curva de supervivencia para el grupo tratado con el tratamiento nuevo (línea roja y discontinua) está por encima de la supervivencia esperada en el grupo control (tratamiento estándar).

También podemos comparar la estimación de los percentiles (25, 50 y 75) del tiempo de supervivencia de ambos grupos:

```
> quantile(.Survfit, quantiles=c(.25,.5,.75))
, , Stratum = trata=Control

      Survival Probability
Estimate      0.25 0.5 0.75
lower 0.95 CL    8  4  2
quantile         12  8  4
upper 0.95 CL   NA 12  8

, , Stratum = trata=Tratamiento

      Survival Probability
Estimate      0.25 0.5 0.75
lower 0.95 CL   23 16  6
quantile         NA 23 13
upper 0.95 CL   NA NA  NA
```

en los que podemos observar que claramente el grupo “Tratamiento” es superior al grupo “Control”.

Ejemplo 1: Regresión de Cox con R-Commander

Banco de datos: **Mieloma.rda**

48 pacientes con mieloma múltiple de los que se conoce:

Tiempo: tiempo de supervivencia (en meses)

Estado: 1=fallecido, 0=dato censurado

Edad: Edad al diagnóstico (años)

Sexo: Sexo (1-Hombre, 2-Mujer)

US: Urea en sangre

HB: Hemoglobina

Información sobre pacientes con mieloma múltiple de los que se conoce: tiempo de supervivencia en meses, logaritmo de urea en sangre, hemoglobina, calcio, sexo y edad al diagnóstico.

Objetivo: Valorar la influencia de las covariables disponibles en el tiempo de supervivencia)

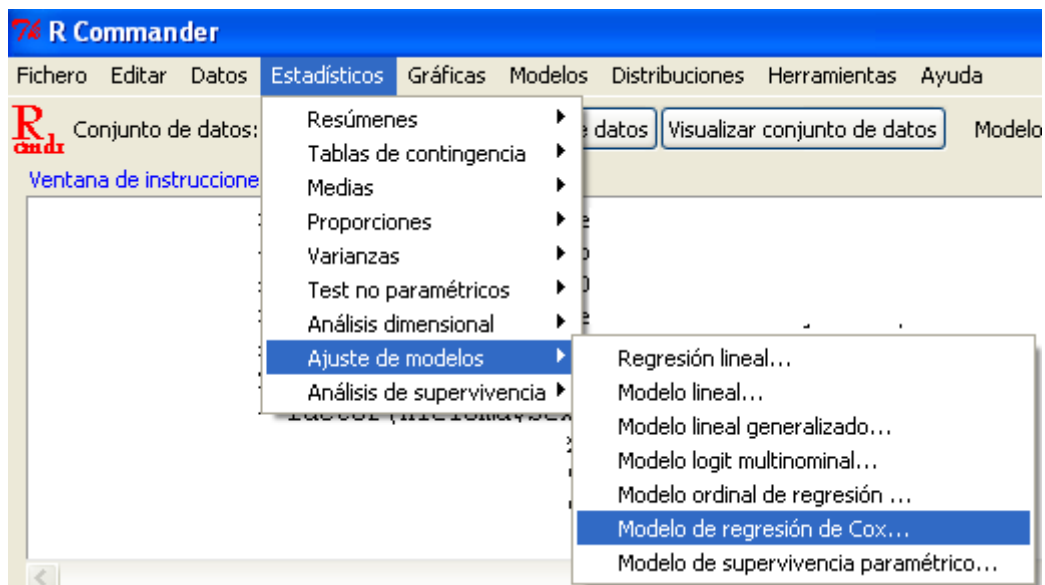
□ Paso 1: Cargar banco de datos: RecaidaLeucemia.rda (tal y como se cargan habitualmente)

□ Paso 2: Indicar al programa cuál es la variable que guarda el tiempo y cuál es la que guarda la censura.

Indicar que la variable “Tiempo” recoge el tiempo de supervivencia desde el diagnóstico hasta la muerte y la variable “Estado” es la que recoge el evento o censura de cada individuo

□ Paso 3: Regresión de Cox.

Acceder en el menú a la opción: Estadísticos> Ajuste de modelos> Modelo de regresión de Cox



Aparecerá la ventana:

Modelo de regresión-Cox

Introduzca el nombre del modelo: CoxModel.5

Tiempo o tiempos de comienzo/final (seleccione uno o dos): Edad, Estado, HB, Tiempo

Indicador de evento (seleccione uno): Edad, Estado, HB, Tiempo

Estratos (seleccione cero o más): Sexo

Grupos (opcional): Sexo

Método para empates: Efron (seleccionado), Breslow, Exacto

Errores Estándar Robustos: Defecto (seleccionado), Sí, No

Variables (doble clic para enviar a la fórmula): HB, Sexo [factor], Tiempo, US

~ Edad +Sexo +HB +US |

Expresión de selección: <todos los casos válidos>

Aceptar Cancelar Ayuda

El resto de opciones son las mismas que en el caso de la estimación de la Curva de Supervivencia

Indicamos el modelo deseado, de la misma forma que lo hacíamos en Regresión lineal o Regresión Logística.

Tras indicar qué variables queremos introducir en el modelo, pulsamos Aceptar y el programa nos devuelve los resultados que se muestran a continuación:

RR (Riesgo relativo)

	coef	exp(coef)	se(coef)	z	Pr(> z)
Edad	-0.022818	0.977441	0.027423	-0.832	0.40538
Sexo[T.Mujer]	-0.266016	0.766427	0.376437	-0.707	0.47977
HB	-0.167713	0.845597	0.065015	-2.580	0.00989 **
US	0.018808	1.018986	0.005892	3.192	0.00141 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Edad	0.9774	1.0231	0.9263	1.0314
Sexo[T.Mujer]	0.7664	1.3048	0.3665	1.6029
HB	0.8456	1.1826	0.7444	0.9605
US	1.0190	0.9814	1.0073	1.0308

Rsquare= 0.277 (max possible= 0.989)
Likelihood ratio test= 15.55 on 4 df, p=0.003691
Wald test = 18.27 on 4 df, p=0.001092
Score (logrank) test = 22.77 on 4 df, p=0.0001408

p-valores sobre la significatividad de cada variable sobre el riesgo de muerte (o sobre la supervivencia)

Intervalos de confianza para el RR

Crerios sobre la bondad de ajuste del modelo

Inversa del RR, útil para interpretar los RR menores a 1 (correspondientes a factores protectores)

A partir de estos resultados podemos obtener la estimación del Riesgo relativo para cada covariable, su correspondiente intervalo de confianza y el p-valor sobre su significatividad (¿Es realmente diferente del valor 1?)

Por otro lado, obtenemos la inversa del RR para facilitar la interpretación en el caso de valores del RR inferiores a 1.

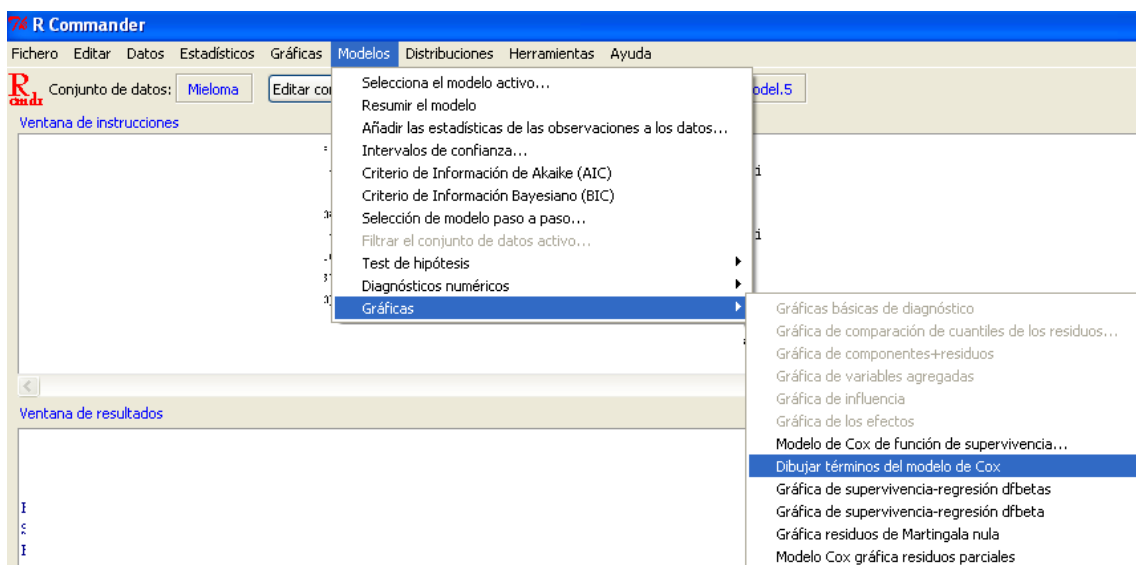
- Edad: La estimación del RR respecto a la edad es 0.977441, con un intervalo de confianza al 95% (0.9263, 1.0314). Como la estimación puntual es inferior a 1, indica que por cada año más de edad disminuye el riesgo de morir, concretamente los que tienen un año menos tienen 1.0231 veces más probabilidad de morir que los que tienen un año más. Observamos que el p-valor es 0.40538 (mayor que 0.05), por lo que la relación entre la edad y la supervivencia estimada no es estadísticamente significativa (hecho que también nos indica el intervalo de confianza, pues contiene el valor 1).
- Sexo: El RR de mujeres respecto a hombres estimado es 0.766427, por lo que las mujeres tienen menos riesgo de morir que los hombres (los hombres tienen 1.3048 veces más riesgo que las mujeres). El intervalo de confianza es (0.3665,1.6029). Este intervalo contiene a 1, por lo que

la relación no es significativa (hecho que también indica el p-valor =0.47977 mayor que 0.05).

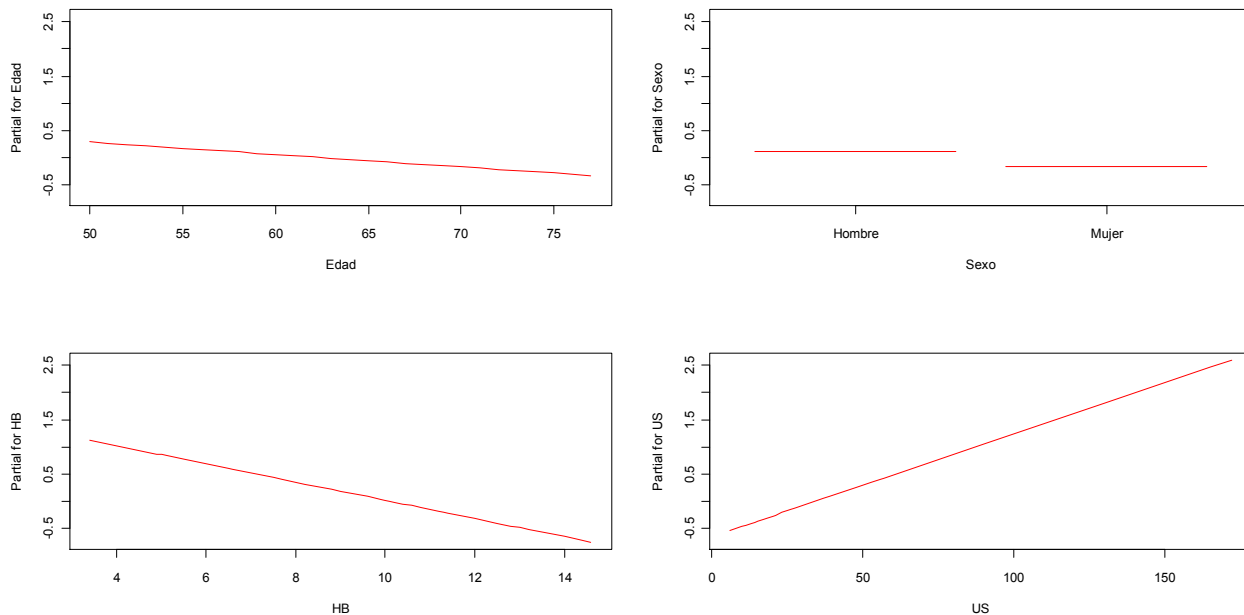
- Hemoglobina: La estimación del RR por cada unidad que aumenta el nivel de Hb es 0.845597 (es decir, por cada unidad menos en el nivel de Hb se multiplica por 1.1826 la probabilidad de morir). El intervalo de confianza para este RR es (0.7444,0.9605) y como podemos observar no contiene el valor 1 y por tanto refleja una relación significativa. Este hecho lo confirma el p-valor =0.00989 (menor que 0.05).
- Urea en sangre: El RR estimado es 1.0189, con un intervalo de confianza (1.0073,1.0308). Por cada unidad que aumenta el nivel de urea en sangre se multiplica por 1.0189 la probabilidad de morir. Esta relación si es significativa, pues el intervalo de confianza no contiene al 1 (y el p-valor es 0.00141, también menor que 0.05).

Así, podemos concluir que no parece haber influencia de la edad y el sexo en la supervivencia tras ser diagnosticado de esta enfermedad, mientras que tanto el menor nivel de hemoglobina y el mayor nivel de urea en sangre afectan negativamente en el tiempo de supervivencia (o lo que es lo mismo, incrementan el riesgo de morir).

Para obtener una estimación del efecto de estas variables podemos solicitar la grafica disponible en la opción del menú:Modelos> Gráficas> Dibujar términos del modelo de Cox



La representación gráfica que se obtiene es la que se muestra a continuación:



En esta representación gráfica se puede observar las relaciones que hemos comentado a la vista de los resultados.

ACTIVIDAD:

El fichero **laringe.rda** contiene los datos de un estudio realizado en Holanda con 90 hombres diagnosticados de cáncer de laringe. Tiene la siguiente estructura:

tiempo: Tiempo desde el tratamiento hasta la muerte (o final de seguimiento) en meses

censura: Indica 1=muerto, 0=vivo (censura)

edad: Edad en el momento del diagnóstico (años)

año: Año de diagnóstico

estadio: Estadio de la enfermedad (clasificada en 4 niveles o categorías: de 1 a 4)

ACTIVIDADES:

- 1) Estima la curva de supervivencia por el método de Kaplan-Meier, obtén la gráfica con intervalos de confianza para la misma y comenta los resultados
- 2) Compara las cuatro curvas de supervivencia de los diferentes grupos que definen los diferentes estadios de la enfermedad. ¿Existen diferencias significativas entre ellas? Obtén una representación gráfica de las cuatro curvas de supervivencia (Kaplan-Meier) y comenta los resultados.
- 3) Introduce las variables en un modelo de regresión de Cox y comenta las conclusiones que de él se deriven (acompaña los resultados de la gráfica de los términos del modelo de Cox)