

## CASO PRÁCTICO: ANÁLISIS DE SUPERVIVENCIA.

Tal y como hemos descrito en el tema, este tipo de análisis lo podemos hacer con Rstudio o con Rcomander. Vamos a comenzar con este último por su facilidad.

Para este tipo de análisis con Rcmdr necesitamos un paquete específico: 'RcmdrPlugin.survival'. Ya describimos como instalar paquetes en R, [Paquetes → instalar paquete\(s\)...](#), buscamos en el repositorio de Spain (Madrid) [https] y buscamos el paquete 'RcmdrPlugin.survival'. Una vez instalado hay que cargarlo igual que con Rcmdr, pero en este caso cargando el paquete ya comentado.

Como podéis observar, nos aparece la misma ventana de Rcmdr. La única diferencia es que si seleccionáis el menú estadísticos, veréis que al final aparece "Análisis de supervivencia"

Empezamos como en otros casos cargando el archivo que queremos estudiar. En este caso se trata del archivo ovario.xlsx con tres variables: las características de este tipo de estudios ("tiempo" y "censura") y una variable cualitativa, "tratamiento" (placebo y tratadas).

Ya sabéis como cargar el archivo: [datos → importar datos desde Excel](#). Nos aparece una ventana. En ella ponemos el nombre del archivo y marcamos las opciones que nos interesen. Damos a "aceptar" y nos aparece el explorador de "Windows". Para buscar la ubicación de nuestro archivo lo seleccionamos y listo para trabajar. Veréis que el nombre que hemos dado a nuestro archivo está en la ventana de "Conjunto de datos".

Podemos hacer un estudio descriptivo de las variables si nos interesa, ya sabríais como hacerlo.

Para hacer el estudio de supervivencia nos vamos a la ventana de estadísticos y seleccionamos "Análisis de supervivencia", y dentro de ésta "Estimación de la función de supervivencia". Nos aparece la siguiente ventana:

Función de supervivencia

Datos Opciones

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

Indicador de evento (seleccione uno o más)  
censura  
tiempo

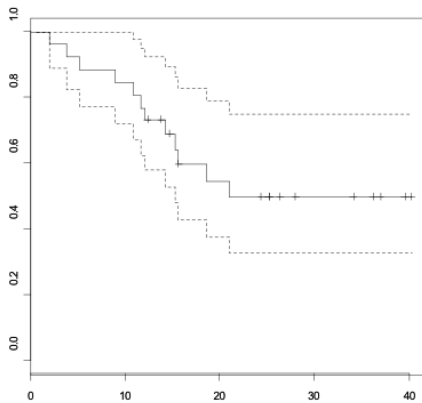
Estratos (seleccione cero o más)  
tratamiento

Tipo de censura  
 Defecto  
 Derecha  
 Izquierda  
 Intervalo  
 Intervalo tipo 2  
 Counting process

Expresión de selección  
<todos los casos válid

Ayuda Reiniciar Aceptar Cancelar Aplicar

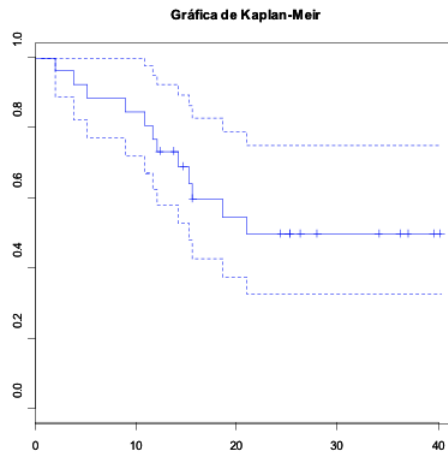
Marcamos la variable "tiempo", a la derecha y "censura" a la izquierda y le damos a aceptar:



Obtenemos la gráfica de Kaplan-Meier de supervivencia, que ya hemos explicado en el tema.

La gráfica la podemos personalizar un poco como ya hemos descrito en otros temas. Bastaría con editar la orden de R e introducir los comando de color (col=), título de los ejes (xlab=, y lab=) o un título (main=)

La gráfica personalizada podría ser la siguiente:



Ahora nos interesa ver las gráficas de las mujeres que han tomado tratamiento y las que han tomado placebo y ver como son las gráficas. Hacemos lo mismo que hemos hecho para la gráfica de supervivencia, pero en la ventana donde hemos marcado "tiempo" y "censura", marcamos, además, la variable "tratamiento". La gráfica que se obtiene es la siguiente:

R Función de supervivencia

Datos Opciones

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

censura tiempo

Indicador de evento (seleccione uno o más)

censura tiempo

Estratos (seleccione cero o más)

tratamiento

Expresión de selección

Tipo de censura

Defecto

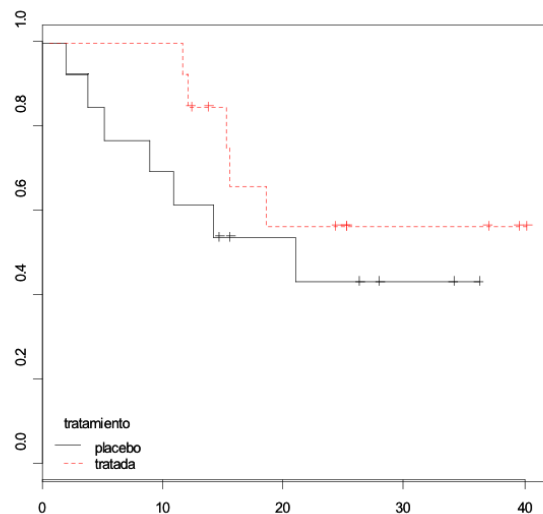
Derecha

Izquierda

Intervalo

Intervalo tipo 2

Counting process



Nos quedaría, por último, realizar el test de LogRank para ver si realmente las curvas son iguales o diferentes. Para ello nos vamos a **Estadísticos → Análisis de supervivencia → Comparación de las funciones de Supervivencia**

Nos sale una ventana muy similar a la que hemos abierto para hacer las gráficas. Marcamos las tres variables: "tiempo", "censura" y "tratamiento". Aceptamos y los resultados son los siguientes:

```
survdif(formula = Surv(tiempo, censura) ~ tratamiento, data = ovario, rho = 0)
n=26, 111 observations deleted due to missingness.
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
tratamiento=placebo	13	7	5.23	0.596	1.06
tratamiento=tratada	13	5	6.77	0.461	1.06

```
Chisq= 1.1 on 1 degrees of freedom, p= 0.303
```

Nuestra hipótesis nula sería la igualdad de las funciones de supervivencia, como  $p$  es muy grande ( $>0.05$ ), no podemos rechazarla y, por tanto, podemos decir que no hay diferencia en las curvas.

Como ya hemos comentado, trabajar con R directamente o con la ayuda de Rstudio nos da más posibilidades. Vamos a describir los comandos que necesitamos para estudiar la función de supervivencia, así como los datos de incidencia acumulativa y otros que hemos descrito en el tema.

Comenzamos instalando el paquete: "Survival". Como ya hemos descrito, nos vamos a la ventana inferior derecha, en "Packages", nos vamos a install, escribimos "survival" y marcamos la casilla de "install dependencies".

Vamos a trabajar con un archivo diferente del que hemos empleado para trabajar con Rcmdr: "bmt.csv". Es un estudio de supervivencia tras trasplante de médula ósea. Lo cargamos desde Rstudio seleccionando archivos csv. Buscamos la ubicación del archivo y lo importamos.

Comenzamos el estudio cargando el paquete y el archivo. Los comandos son:

```
library(survival)
attach(bmt)
```

Con ello hemos instalado el paquete survival y hemos cargado el archivo. Ahora nos interesa ordenar el archivo en función del tiempo. La orden es la siguiente:

```
bmt<-bmt[order(tiempo),]
```

Con esto hemos creado un objeto que es el mismo archivo "bmt", pero ordenado en función del tiempo. Si queremos ver los datos, podemos ver los 5 primeros registros con la orden:

```
Head(bmt)
```

R nos muestra los siguientes datos:

x1	id	grupo	tiempo	censura	t_recu	recu_plaq	edadp	edadd	fab	imtx	sexop	sexod
1	35	AML- altoriesgo	1	1	1	no	42	48	0	si	hombre	hombre
2	108	AML- altoriesgo	2	1	2	no	20	19	1	no	hombre	hombre
3	84	AML- altoriesgo	10	1	10	no	34	54	0	si	hombre	mujer
4	129	AML- altoriesgo	16	1	16	no	27	36	0	so	mujer	mujer
5	114	AML- altoriesgo	32	1	16	si	36	39	1	so	hombre	hombre
6	87	AML- altoriesgo	35	1	35	no	23	25	0	si	mujer	hombre

He tabulado los datos y he acertado algún nombre de variable para poder mostrarlos en el texto y, como podéis observar, están ordenados en función del tiempo.

El archivo tiene 13 variables de las cuales las dos primeras no tienen interés.

Comenzamos a describir cómo realizar el estudio de supervivencia. Para ello, vamos a mostrar cómo hacer la tabla que usamos en el tema para describir como se construye la curva:

Función de supervivencia:

```
resp<-Surv(tiempo,censura)
fit<-survfit(resp~1,data=bmt)
```

Si queremos ver los datos:

```
Summary(fit)
```

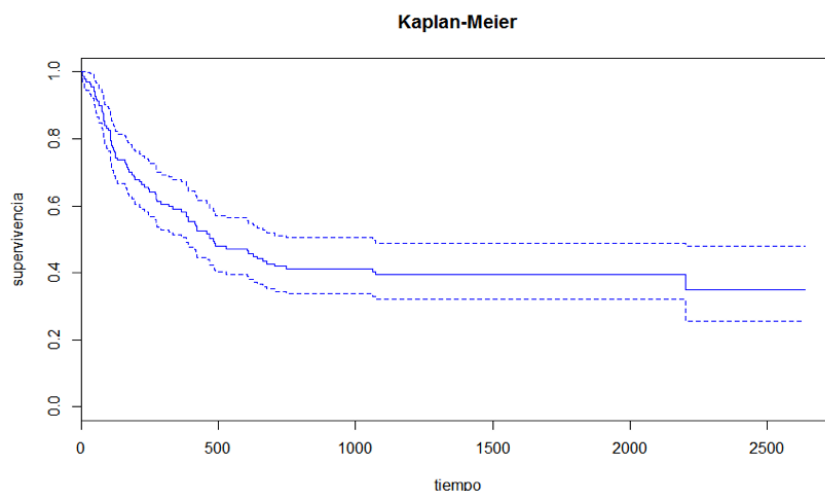
Estos son:

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	137	1	0.993	0.00727	0.979	1.000
2	136	1	0.985	0.01025	0.966	1.000
10	135	1	0.978	0.01250	0.954	1.000
16	134	1	0.971	0.01438	0.943	0.999
32	133	1	0.964	0.01602	0.933	0.995
35	132	1	0.956	0.01748	0.923	0.991
47	131	2	0.942	0.02003	0.903	0.982
48	129	2	0.927	0.02222	0.884	0.972
53	127	1	0.920	0.02322	0.875	0.966
55	126	1	0.912	0.02415	0.866	0.961
63	125	1	0.905	0.02504	0.857	0.956
64	124	1	0.898	0.02588	0.848	0.950
74	123	2	0.883	0.02744	0.831	0.939
76	121	1	0.876	0.02817	0.822	0.933
79	120	1	0.869	0.02886	0.814	0.927

En el cuadro os pongo el principio de la tabla y la que explicamos en el tema para explicar el estudio de supervivencia. Hay 137 individuos como podéis ver.

Si queremos dibujar la curva la orden es la siguiente:

```
plot(fit,xlab="tiempo",ylab="supervivencia", main="Kaplan-Meier", col="blue")
```

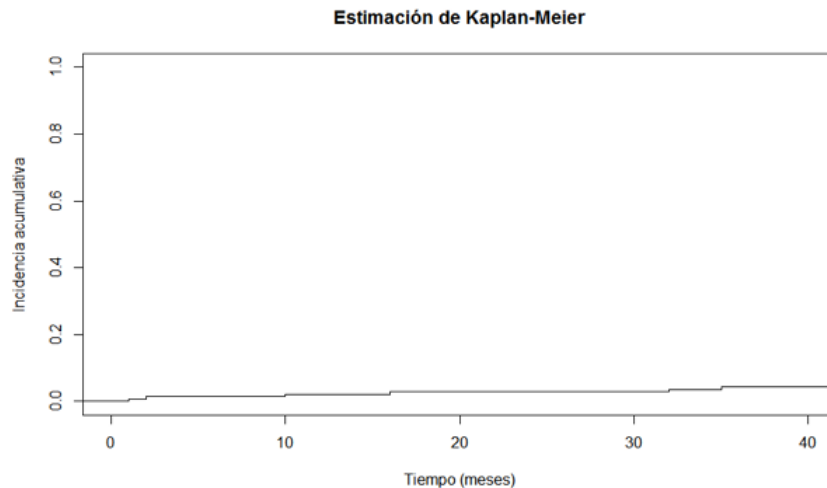


Esta es la curva que se obtiene con la orden que hemos ejecutado. Ahora nos puede interesar calcular la incidencia acumulativa. La orden es la siguiente:

```
#Estimación de la función de incidencia acumulativa
super<-fit$surv
ia<-1-super
print(ia)
ia.steps <- stepfun(fit$time, c(0, ia))
plot(ia.steps, do.points = FALSE, xlim = c(0, 40), ylim = c(0,1),
     ylab = "Incidencia acumulativa", xlab = "Tiempo (meses)", main =
     "Estimación de Kaplan-Meier")
```

Con esta orden podemos ver la incidencia acumulativa y la gráfica correspondiente:

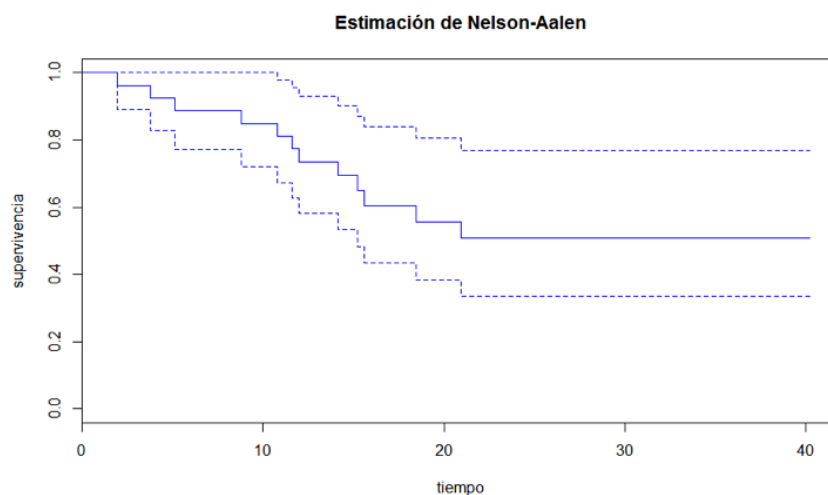
```
> print(ia)
 [1] 0.00729927 0.01459854 0.02189781 0.02919708 0.03649635 0.04379562 0.
05839416 0.07299270 0.08029197 0.08759124
 [11] 0.09489051 0.10218978 0.11678832 0.12408759 0.13138686 0.14598540 0.
15328467 0.16058394 0.16788321 0.17518248
 [21] 0.18248175 0.19708029 0.20437956 0.21167883 0.21897810 0.22627737 0.
23357664 0.24087591 0.25547445 0.26277372
 [31] 0.27007299 0.27737226 0.28467153 0.29197080 0.29927007 0.30656934 0.
31386861 0.32116788 0.32846715 0.33576642
 [41] 0.33576642 0.34314680 0.35052717 0.35790754 0.36528792 0.37266829 0.
38004866 0.38742903 0.39480941 0.40218978
 [51] 0.40957015 0.41695053 0.42433090 0.43171127 0.44647202 0.45385239 0.
46123277 0.46861314 0.47599351 0.48337388
 [61] 0.49075426 0.49813463 0.50551500 0.51289538 0.52027575 0.52765612 0.
52765612 0.53515365 0.54265117 0.55014869
 [71] 0.55764621 0.56514373 0.57264125 0.58013878 0.58763630 0.58763630 0.
58763630 0.58763630 0.58763630 0.58763630
 [81] 0.58763630 0.58763630 0.58763630 0.59640999 0.60518369 0.60518369 0.
60518369 0.60518369 0.60518369 0.60518369
 [91] 0.60518369 0.60518369 0.60518369 0.60518369 0.60518369 0.60518369 0.
60518369 0.60518369 0.60518369 0.60518369
 [101] 0.60518369 0.60518369 0.60518369 0.60518369 0.60518369 0.60518369 0.
```



En el texto del tema comentamos que hay otras funciones de supervivencia, y en concreto hablamos de la función de Nelson Aalen. La orden es la siguiente:

```
fit_na<-survfit(resp~1, type="fleming")
summary(fit_na)
plot(fit_na,main="Estimación de Nelson
Aalen",ylab="supervivencia",xlab="tiempo", col="blue")
```

La orden "summary" nos presenta una tabla similar a la de Kaplan-Meier y la gráfica de supervivencia también es muy similar



Vamos a representar ahora la curva en función de la variable "grupo" (tipo de leucemia). Las órdenes de los comandos serían las siguientes:

```
resp<-Surv(tiempo,censura)
fit1<-survfit(resp~grupo,data=bmt)
summary(fit1)
plot(fit1,xlab="tiempo",ylab="supervivencia",lty=c(1,2), col=c("blue","red","green"))
legend(locator(1),legend=c("ALL","AML-alto riesgo","AML-bajo
riesgo"),lty=c(1,2),bty="n",col=c("blue","red","green"))
```

Por último, nos queda realizar el test de LogRank para ver si las curvas son iguales o diferentes. La orden es la siguiente:

```
survdif(Surv(tiempo, censura) ~ grupo,data=bmt)
```

Una vez ejecutada la orden R nos da el resultado del test:

```
survdif(formula = Surv(tiempo, censura) ~ grupo, data = bmt)

```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
grupo=ALL	38	24	21.9	0.211	0.289
grupo=AML-alto riesgo	45	34	21.2	7.756	10.529
grupo=AML-bajo riesgo	54	25	40.0	5.604	11.012

En este caso la  $p$  es menor de 0,05, es decir muy pequeña, por lo que puedo afirmar con una probabilidad de error de 0.001 que las tres curvas son diferentes.

Con esto finalizamos el estudio de las curvas de supervivencia. Ya hemos explicado en el tema que el modelo multivariante de este tipo de estudios es el modelo de Cox.

Lo vamos a estimar en primer lugar en un modelo univariante. La primera variable que vamos a estudiar es la edad del paciente. La orden es la siguiente:

```
fit3<-coxph(resp~edadp,data=bmt)
summary(fit3)
```

El resultado del modelo es el siguiente:

```
coxph(formula = resp ~ edadp, data = bmt2)

```

n= 137, number of events= 83

	coef	exp(coef)	se(coef)	z	Pr(> z )
edadp	0.01122	1.01129	0.01171	0.959	0.338

	exp(coef)	exp(-coef)	lower .95	upper .95
edadp	1.011	0.9888	0.9883	1.035

Concordance= 0.544 (se = 0.033 )  
 Rsquare= 0.007 (max possible= 0.996 )  
 Likelihood ratio test= 0.9 on 1 df, p=0.3419  
 Wald test = 0.92 on 1 df, p=0.3377

R nos muestra el coeficiente de la edad del paciente, que recordar que era la exponencial de lo que aumenta el HR por cada unidad de aumento de la edad. R también nos da esta exponencial, el error estándar y el estadístico Z. Además, nos da el intervalo de confianza de la exponencial de la edad del paciente.

Vamos ahora a introducir más variables en el modelo. La orden es la siguiente:

```
coxph(formula = Surv(tiempo,censura)~grupo+edadp+sexop,data = bmt)
```

```
n= 137, number of events= 83
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
grupoAML-alto riesgo	0.274397	1.315737	0.290573	0.944	0.3450
grupoAML-bajo riesgo	-0.670123	0.511646	0.299360	-2.239	0.0252 *
edadp	0.008086	1.008118	0.012667	0.638	0.5232
sexopmujer	0.269199	1.308915	0.225839	1.192	0.2333

```
---
```

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

	exp(coef)	exp(-coef)	lower .95	upper .95
grupoAML-alto riesgo	1.3157	0.7600	0.7444	2.325
grupoAML-bajo riesgo	0.5116	1.9545	0.2845	0.920
edadp	1.0081	0.9919	0.9834	1.033
sexopmujer	1.3089	0.7640	0.8408	2.038

```
Concordance= 0.636 (se = 0.033 )
```

```
Rsquare= 0.106 (max possible= 0.996 )
```

```
Likelihood ratio test= 15.35 on 4 df, p=0.004021
```

```
Wald test = 14.92 on 4 df, p=0.004868
```

```
Score (logrank) test = 15.69 on 4 df, p=0.003464
```

Los datos son los ya explicados pero, al haber más variables, nos da los datos para cada una de las variables estudiadas. Fijaos que la variable "grupo" es una variable *dummy*, por lo que compara con la categoría de referencia, que en este caso es "ALL". El modelo de cox también admite estudiar el término de interacción explicado en el modelo multivariante. Se hace sustituyendo el signo "+" por el de multiplicar. La interpretación es la misma que se explicó en el modelo multivariante.

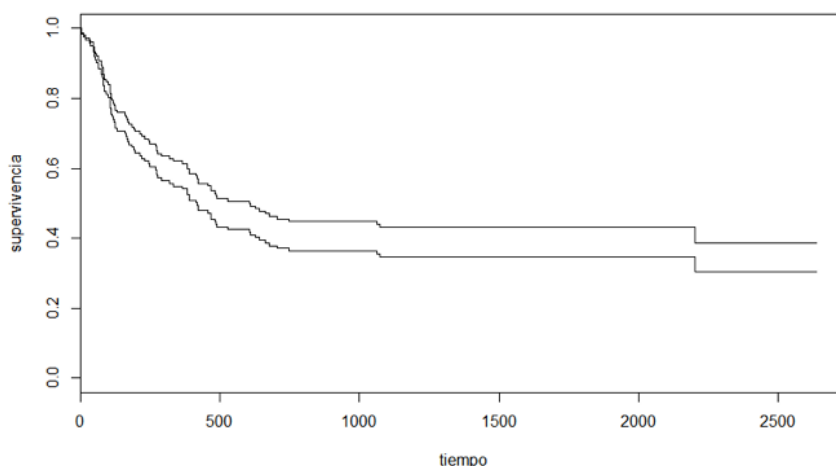
La representación gráfica del modelo de Cox, debido a su carácter multivariante, es difícil de hacer cuando hay muchas variables o estas son numéricas. Vamos a ver un ejemplo con una sola variable dicotómica:

La orden es la siguiente:

```
fit3<-coxph(Surv(tiempo,censura)~sexop,data=bmt)
summary(fit3)
plot(survfit(fit3, newdata=bmt), lty=c(1,2),mark.time=FALSE,
      xlab="tiempo",ylab="supervivencia")
```

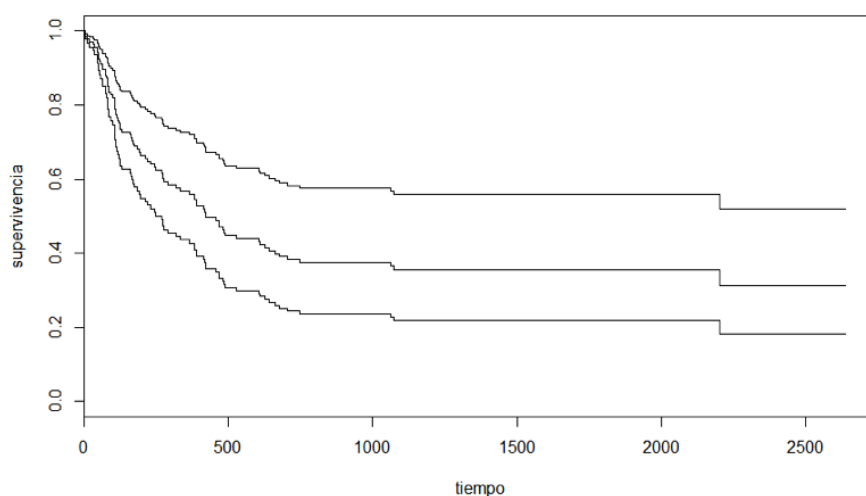
La gráfica obtenida es la siguiente:





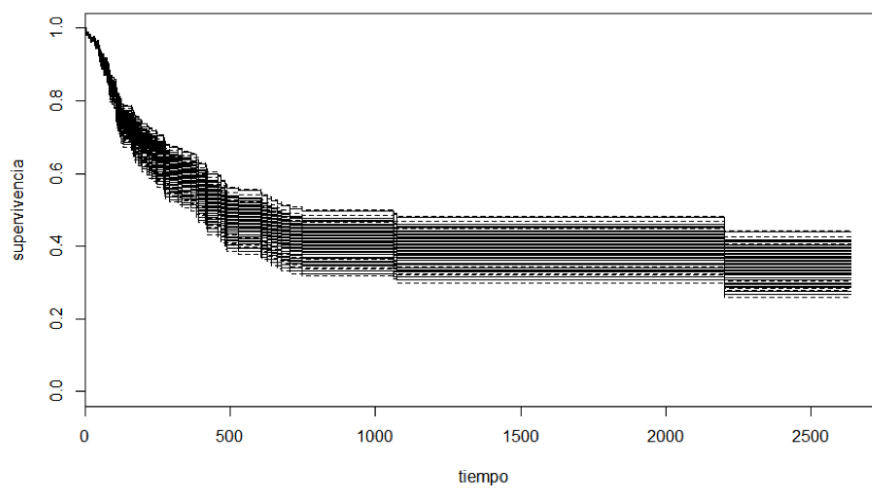
Vamos a hacer una gráfica ahora con una variable categórica con más de dos categorías. En nuestro caso, la variable "grupo":

```
fit5<-coxph(Surv(tiempo,censura)~grupo,data=bmt)
summary(fit5)
plot(survfit(fit5, newdata=bmt), lty=c(1,2),mark.time=FALSE,
xlab="tiempo",ylab="supervivencia")
```



Esta gráfica ya la hemos hecho con el método de Kaplan Meier. Vamos ahora a dibujar la gráfica de una variable cuantitativa. En este caso, al igual que en las variables cualitativas, se imprime un dato por cada valor de la variable resultando la gráfica siguiente:

```
fit6<-coxph(Surv(tiempo,censura)~edad,data=bmt)
summary(fit6)
plot(survfit(fit3, newdata=bmt), lty=c(1,2),mark.time=FALSE,
xlab="tiempo",ylab="supervivencia")
```



Como veís, en este caso es difícil de interpretar la gráfica, ya que nos dibuja una línea por cada edad del archivo.