

Utiliza la base de datos de R llamada larynx, del paquete KMsurv, y realice lo siguiente:

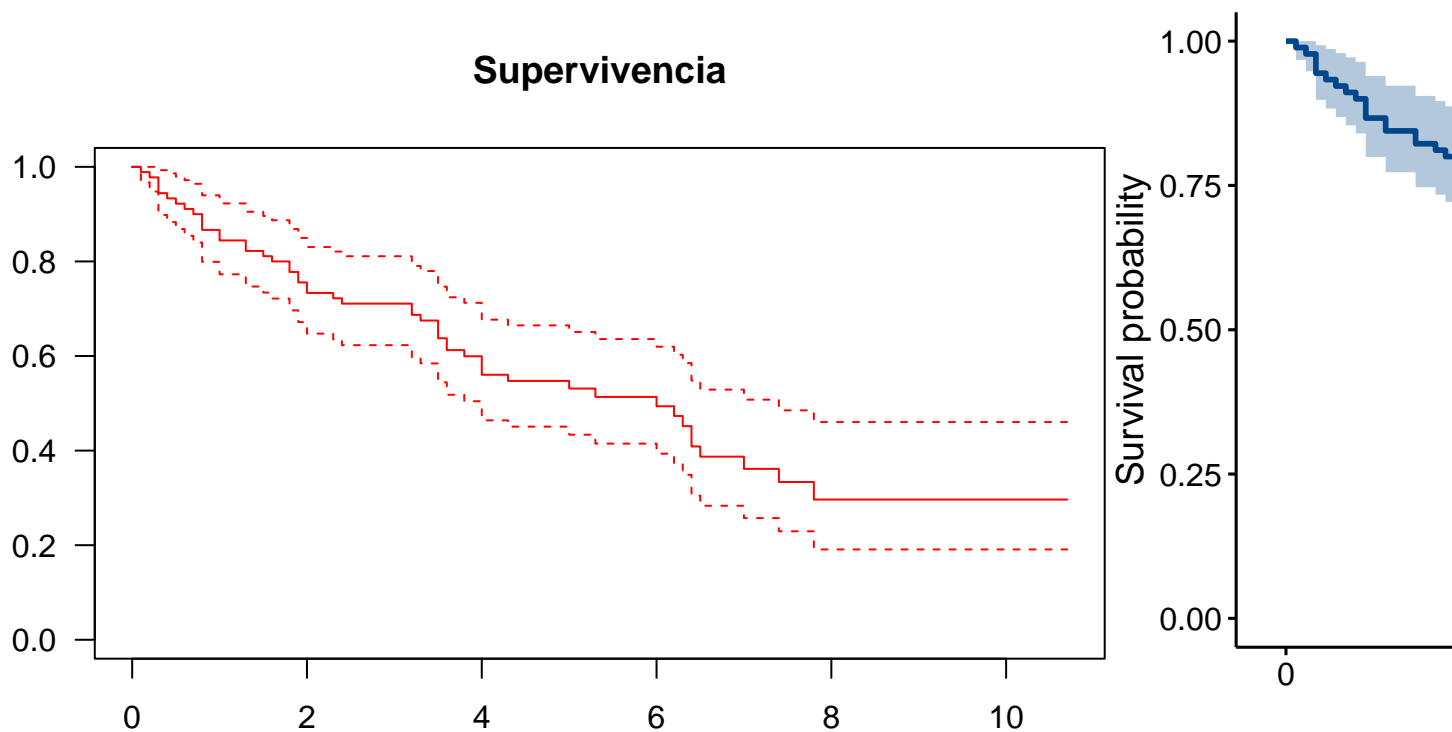
Explique sus resultados.

Análisis descriptivo

1.- Realice un análisis descriptivo sobre el tiempo de supervivencia de los sujetos, además de sus características generales y particulares.

Grafiquemos y revisemos:

```
data("larynx")
Surv_general <- Surv(larynx$time, larynx$delta)
general_fit <- survfit(Surv_general ~ 1)
plot(general_fit, main="Supervivencia", col="red", las=1) #ponlas bonito yisus
ggsurvplot(general_fit, data=larynx, palette = 'lancet', censor.size=4.9, censor.shape=124)
```



Podemos observar que al inicio del estudio, en los primeros dos meses, se tiene una gran cantidad de fallas, seguido por el periodo de 3 a 4 meses y de 6 a 6 y medio meses, aproximadamente; siendo los periodos donde más se concentran las fallas.\

Otro punto a destacar es que no llegamos a 0 en la función de supervivencia, esto porque los últimos datos son censuras, como podemos ver en el dataset (en este contexto es que se encuentran vivos).\

Podemos notar un periodo de estabilidad, donde la función de supervivencia se mantiene constante, en el periodo de aproximadamente 4 meses y medio a poco antes de 6, si bien tenemos fallas, estas son pocas, por lo que se mantiene casi constante en el periodo mencionado, aunque existen muchas fallas (bajo este contexto, ello significa que el sujeto de estudio se encontraba vivo al tiempo de observación).

Al inicio, los intervalos de confianza están bastante cercanos al estimador puntual.\ Los intervalos de confianza se hacen muy grandes hacia el final de la función de supervivencia, pero sesgados hacia la derecha, es decir,

el intervalo de confianza superior tiene una distancia mayor al estimador puntual, respecto del intervalo inferior.\

Dicho lo anterior, se puede intuir que los primeros dos meses son los ‘más difíciles’, puesto que hay muchas fallas en este periodo (son frecuentes y es donde más se acumulan), pero observamos que conforme se acerca a los 8 meses se entra a un periodo “seguro”, donde las fallas son casi nulas, y al tiempo de observación vemos sujetos vivos (fallas) únicamente; teniendo un valor de entre 0.22 y 0.40 aproximadamente, la función de supervivencia con los intervalos de confianza, y de aproximadamente 0.26 con el estimador puntual, por lo que tienen todavía una “Buena” probabilidad de sobrevivir al menos hasta este periodo.

Kaplan-Meier

2.- Con el estimador de Kaplan-Meier para la función de supervivencia $S(t)$, calcule y grafique:

$S(t)$ poblacional

$S(t)$ por estadio de la enfermedad

$S(t)$ por grupos de edad

Identifique las variables que afectan el tiempo de supervivencia. Incluya los intervalos del 95% confianza.

Pruebas Log-Rank con $\alpha = 0.05$

3.- Usando pruebas No paramétricas Log-Rank:

Por estadio

Compare las funciones de supervivencia por estadio, es decir, realice el contraste de hipótesis:

$$H_0 : S_j(t) = S_k(t) \forall t > 0, \forall j, k. \text{ vs } .H_a : S_j(t) \neq S_k(t); \mathbf{p.a} \ t > 0, p.a \ j \neq k.$$

Por grupo de edad

Compare las funciones de supervivencia por grupos de edad