



Tarea 5

Métodos No Paramétricos en Análisis de Supervivencia

Modelos de series de Tiempo y Supervivencia

Profesor: Naranjo Albarrán Lizbeth

Adjuntos: Reyes González Belén

Rivas Godoy Yadira

Integrantes: Cuéllar Chávez Eduardo de Jesús

García Tapia Jesús Eduardo

Miranda Meraz Areli Gissell

Ramírez Maciel José Antonio

Saldaña Morales Ricardo

Grupo: 9249

Fecha: 17/DIC/2021

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

Análisis descriptivo

Ejercicio 1

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

```
## Warning: package 'survminer' was built under R version 4.1.2
```

```
## Warning: package 'ggpubr' was built under R version 4.1.2
```

```
## Warning: package 'ggsci' was built under R version 4.1.2
```

Composición de la base

```
attach(larynx)
head(larynx)
```

```
##   stage time age diagyr delta
## 1     1  0.6  77     76     1
## 2     1  1.3  53     71     1
## 3     1  2.4  45     71     1
## 4     1  2.5  57     78     0
## 5     1  3.2  58     74     1
## 6     1  3.2  51     77     0
```

Esta base de datos es acerca del tiempo de supervivencia en pacientes que padecen cáncer de laringe en distintas etapas.

Tenemos 5 variables y 90 observaciones:

Stage, que representa la etapa en la que se encuentra dicha enfermedad del paciente

Time. que nos dice el tiempo de la observación, en meses

Age, que es edad del paciente cuando se le diagnosticó la enfermedad

Diagyr, que es el año en que se diagnosticó la enfermedad

Delta, que es un indicador tal que es 1 si el paciente murió al tiempo de observación y 0 y se encuentra vivo. En nuestro contexto, la muerte es la falla y el estar vivo la censura

NA y número de observaciones

```
length(larynx$time)
```

```
## [1] 90
```

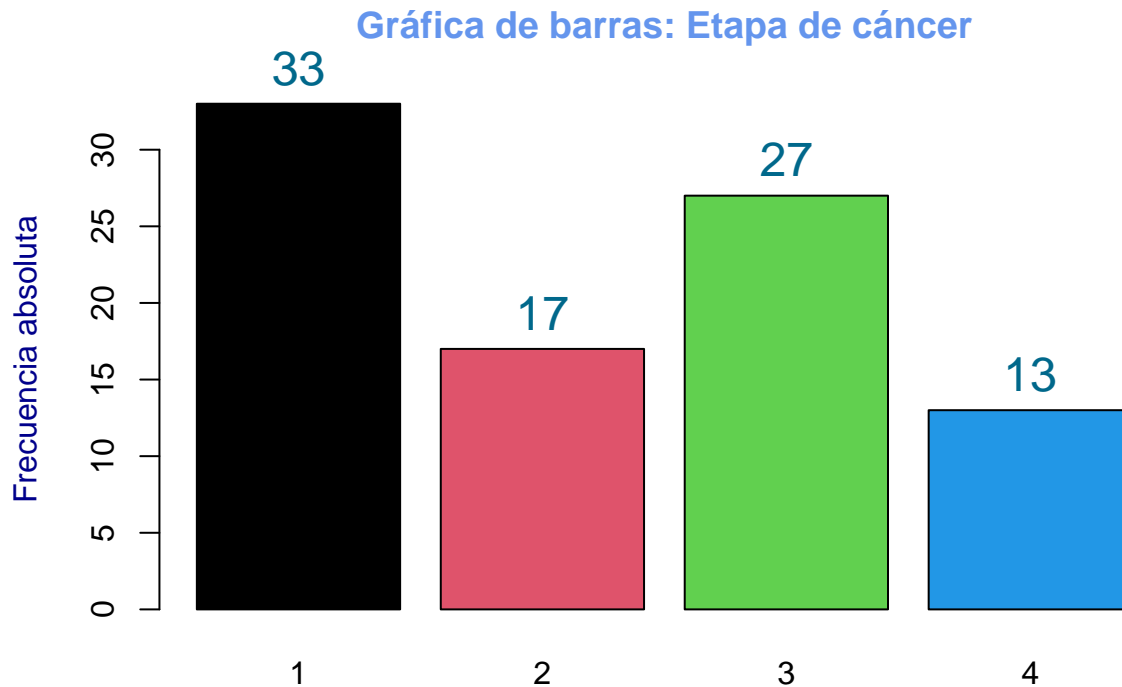
```
sum(is.na.data.frame(larynx))
```

```
## [1] 0
```

Tenemos 5 variables y 90 observaciones, no tenemos ningún NA.

Etapa

```
#Usamos la paleta de lancet porque es una revista de oncología
Edad <- table(stage)
names(Edad) <- unique(larynx$stage)
b <- barplot(Edad,col= palette(pal_lancet("lanonc"))(5)),
      main="Gráfica de barras: Etapa de cáncer", col.main="cornflowerblue",
      ylab="Frecuencia absoluta", col.lab="blue4")
text(x=b,y=c(Edad), labels=c(Edad), pos=3, col="deepskyblue4", cex=1.5, xpd=TRUE)
```



```
#Observaciones
```

```
Edad
```

```
## 1 2 3 4
```

```
## 33 17 27 13
```

```
#Porcentaje
```

```
Edad/.9
```

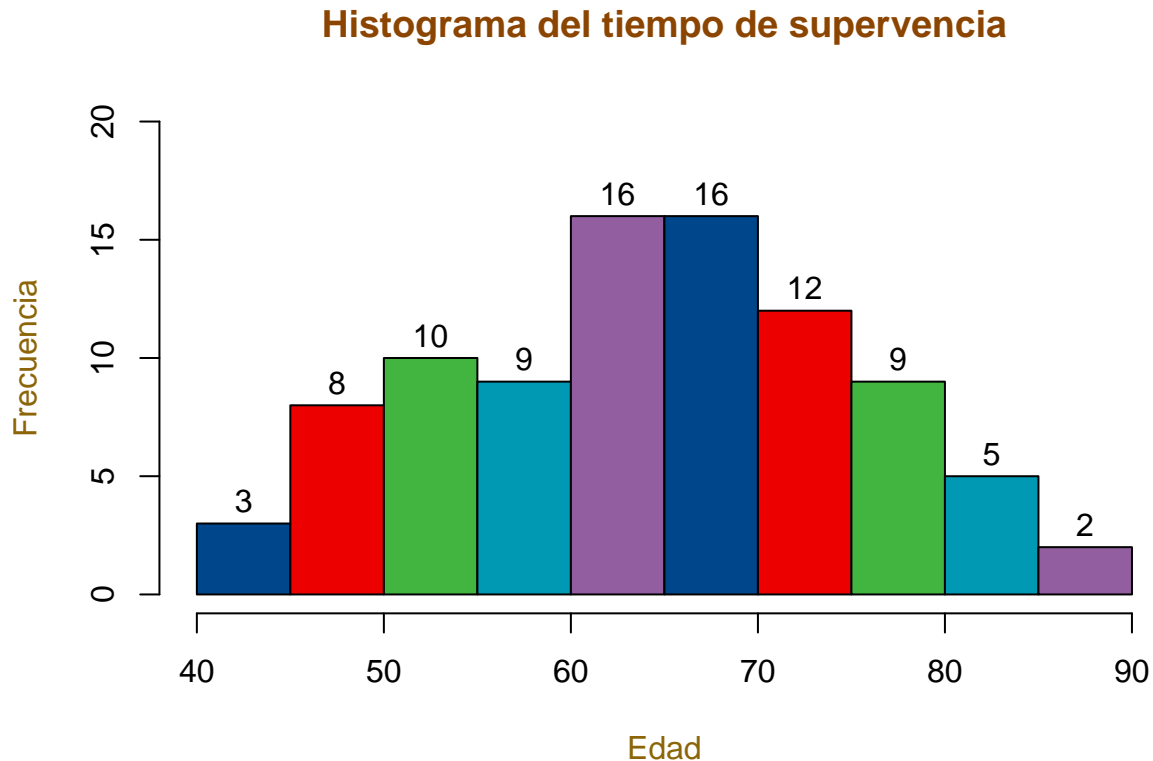
```
## 1 2 3 4
```

```
## 36.66667 18.88889 30.00000 14.44444
```

Podemos ver que la mayoría de las observaciones se concentran en pacientes en etapa 1, que es la inicial, concentrando el 36.66% de las observaciones, después sigue la etapa 3, que tiene 27 observaciones las cuales representan el 30% del total, seguidas de 17 observaciones en etapa 2, representando un 18.8% y al final, como resulta intuitivo, la etapa con menos observaciones es la 4, con 13 que representan el 14.4% del total, esto porque es la etapa más avanzada y grave de la enfermedad, por lo que es entendible que hayan menos pacientes de esta etapa y más de la primera, pues es natural que existan más pacientes en dicho grupo.

Edad

```
edad<-hist(age, probability=F, col=palette(pal_lancet("lanonc")(9)), border="gray0",
  main="Histograma del tiempo de supervencia",col.main="darkorange4",
  ylab = "Frecuencia", col.lab = "darkgoldenrod4",
  xlab = "Edad", labels=TRUE,ylim=c(0,20))
```



#Puntos de corte

```
edad$breaks
```

```
## [1] 40 45 50 55 60 65 70 75 80 85 90
```

#Cantidad

```
edad$counts
```

```
## [1] 3 8 10 9 16 16 12 9 5 2
```

#Porcentaje

```
edad$counts/.9
```

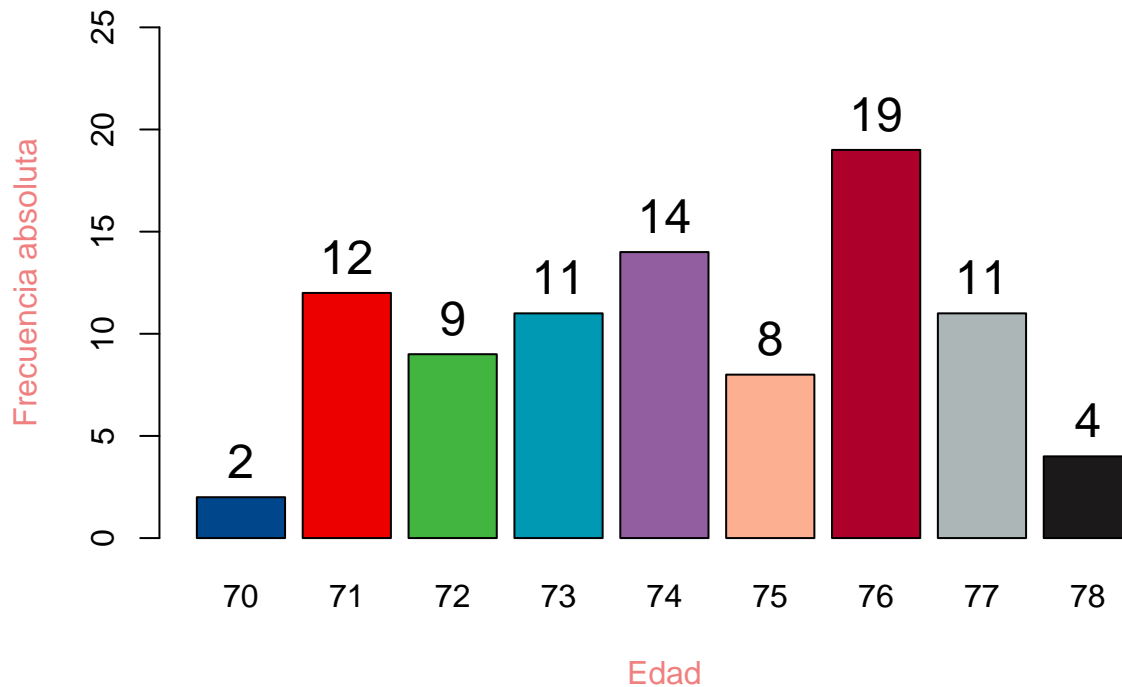
```
## [1] 3.333333 8.888889 11.111111 10.000000 17.777778 17.777778 13.333333
```

```
## [8] 10.000000 5.555556 2.222222
```

Podemos ver que la mayoría de las observaciones se concentran entre 61 y 70 años, teniendo 32 pacientes en dicho grupo de edad, y notamos que mientras se alejan de este grupo de edad, va decreciendo el número de pacientes en ese grupo de edad, parece tener una forma parecida a una distribución simétrica, o acampanada, pero no parece ser una normal, a primera vista, aunque le da un parecido. el grupo más joven y el más viejo tienen la menor concentración de pacientes. ## Año de diagnóstico

```
year <- table(diagyr)
b <- barplot(year,col= palette(pal_lancet("lanonc")(5)),
             main="Gráfica de barras: Etapa de cáncer", col.main="indianred1",
             ylab="Frecuencia absoluta", xlab = "Edad", col.lab="lightcoral",
             ylim=c(0,25))
text(x=b,y=c(year), labels=c(year), pos=3, col="gray0", cex=1.5, xpd=TRUE)
```

Gráfica de barras: Etapa de cáncer



#Observaciones

```
year
```

```
## diagyr
```

```
## 70 71 72 73 74 75 76 77 78
```

```
## 2 12 9 11 14 8 19 11 4
```

#Porcentaje

```
year/.9
```

```
## diagyr
```

```
## 70 71 72 73 74 75 76 77
```

```
## 2.222222 13.333333 10.000000 12.222222 15.555556 8.888889 21.111111 12.222222
```

```
## 78
```

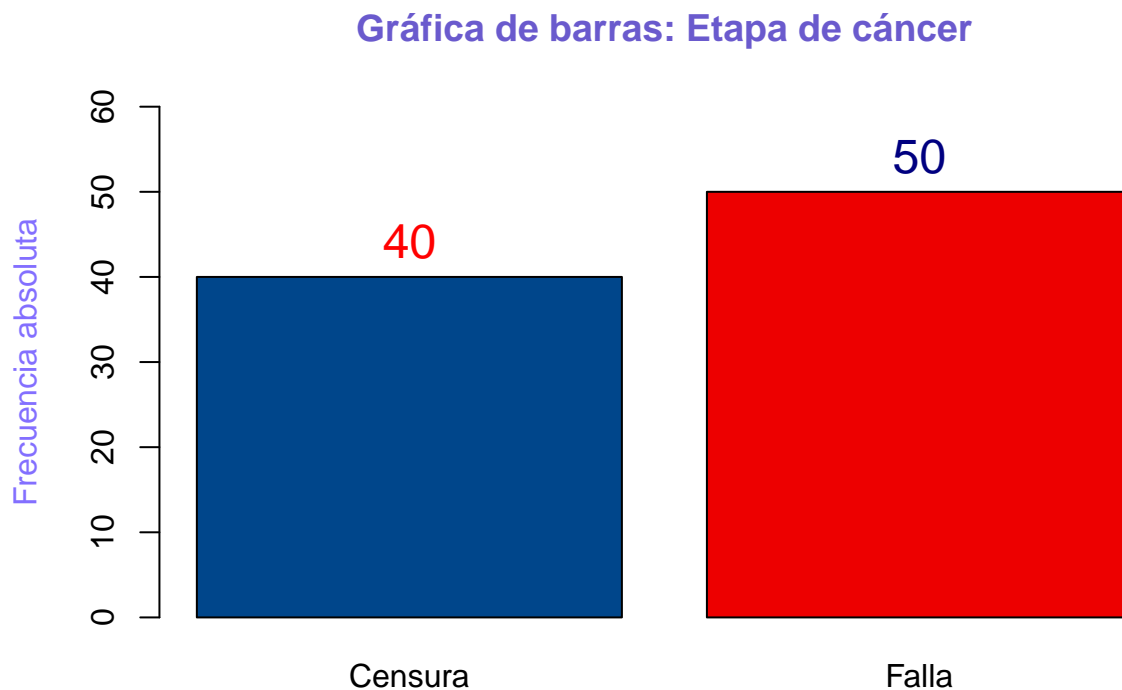
```
## 4.444444
```

Tenemos que el año en el que más personas recibieron su diagnóstico fue en 1976, con 19 personas que representan el 21.11% de los pacientes. Le sigue 1974 con 14 pacientes. Hay casi la misma cantidad de pacientes cuyo diagnóstico fue en 1971, 1973 y 1977, siendo de 12 en el primero y 11 en los otros dos. Después de esto, 1972 y 1975 tienen casi los mismos representantes, siendo 9 y 8 respectivamente. Posteriormente tenemos a 1978, que es el año más reciente, con 4 representantes. Esto puede ser natural ya que, podría costar un

poco de trabajo contactar a personas para el estudio que hayan recibido su diagnóstico muy recientemente, si es que el estudio se realizó el mismo año. Es de esperar que el año con menor pacientes registrados sea 1970, pues es el año más lejano, con solo 2 pacientes.

##Fallas y censuras

```
fallas <- table(delta)
names(fallas)<-c('Censura','Falla')
b <- barplot(fallas,col= palette(pal_lancet("lanonc")(5)),
             main="Gráfica de barras: Etapa de cáncer", col.main="slateblue",
             ylab="Frecuencia absoluta", col.lab="slateblue1",ylim=c(0,60))
text(x=b,y=c(fallas), labels=c(fallas), pos=3, col= c("red", "navy"), cex=1.5,
     xpd=TRUE)
```



```
fallas/.9
```

```
## Censura    Falla
## 44.44444 55.55556
```

Tenemos 50 fallas, que representan el 55.55% de nuestros datos, y 40 censuras, que representan el 44.44% de nuestros datos. Tenemos ligeramente más fallas que censuras en nuestro estudio. ## Función de supervivencia: Tabla actuarial

```
#Elegimos intervalos mensuales
tabla_act_aux<- larynx %>%
  mutate('interval'=floor(larynx$time))
tabla_aux <- table(tabla_act_aux$interval,tabla_act_aux$delta)
intEndpts <- (0:11)
ntotal <- dim(tabla_act_aux)[1]
```

```
cens <- tabla_aux[,1]
fallas <- tabla_aux[,2]
Tabla_Vida_Actuarial<- lifetab(tis = intEndpts, ninit=ntotal, nlost=cens,
                             nevent=fallas)
round(Tabla_Vida_Actuarial,8)
```

##	nsubs	nlost	nrisk	nevent	surv	pdf	hazard	se.surv
## 0-1	90	0	90.0	12	1.0000000	0.13333333	0.14285714	0.00000000
## 1-2	78	0	78.0	10	0.8666667	0.11111111	0.13698630	0.03583226
## 2-3	68	4	66.0	4	0.7555556	0.04579125	0.06250000	0.04530040
## 3-4	60	5	57.5	9	0.7097643	0.11109354	0.16981132	0.04799334
## 4-5	46	8	42.0	4	0.5986708	0.05701626	0.10000000	0.05287161
## 5-6	34	6	31.0	2	0.5416545	0.03494545	0.06666667	0.05498742
## 6-7	26	5	23.5	6	0.5067090	0.12937252	0.29268293	0.05672087
## 7-8	15	4	13.0	3	0.3773365	0.08707766	0.26086957	0.06214054
## 8-9	8	3	6.5	0	0.2902589	0.00000000	0.00000000	0.06503165
## 9-10	5	3	3.5	0	0.2902589	0.00000000	0.00000000	0.06503165
## 10-11	2	2	1.0	0	0.2902589	NA	NA	0.06503165
##	se.pdf	se.hazard						
## 0-1	0.03583226	0.04113397						
## 1-2	0.03312693	0.04321714						
## 2-3	0.02236016	0.03123474						
## 3-4	0.03482957	0.05639938						
## 4-5	0.02758020	0.04993746						
## 5-6	0.02416163	0.04711426						
## 6-7	0.04782305	0.11820092						
## 7-8	0.04636674	0.14932641						
## 8-9	NaN	NaN						
## 9-10	NaN	NaN						
## 10-11	NA	NA						

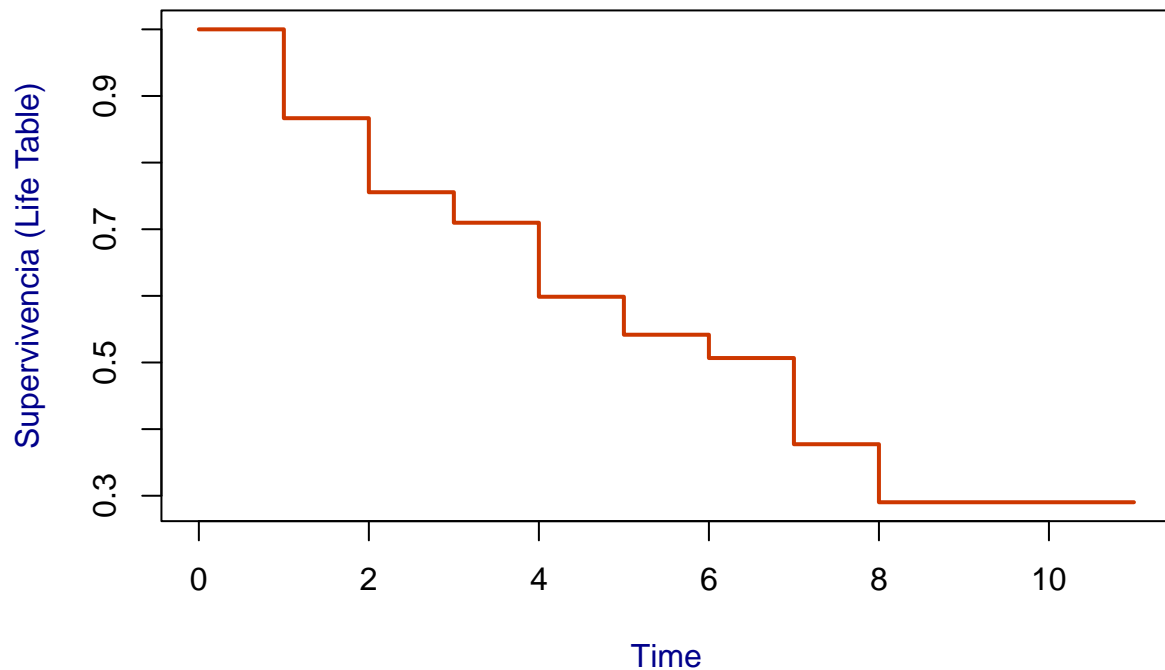
```
names(Tabla_Vida_Actuarial)
```

```
## [1] "nsubs"      "nlost"      "nrisk"      "nevent"     "surv"       "pdf"
## [7] "hazard"     "se.surv"    "se.pdf"     "se.hazard"
```

Graficamos la función de supervivencia:

```
x = rep(intEndpts,rep(2,12))[2:23]
y = rep(Tabla_Vida_Actuarial$surv,rep(2,11))
plot(x, y, type="l", col="orangered3", xlab="Time",
     ylab="Supervivencia (Life Table)", col.lab="darkblue", lwd=2)
title(main="Supervivencia al cáncer", cex=0.6, col.main="darkblue")
```

Supervivencia al cáncer

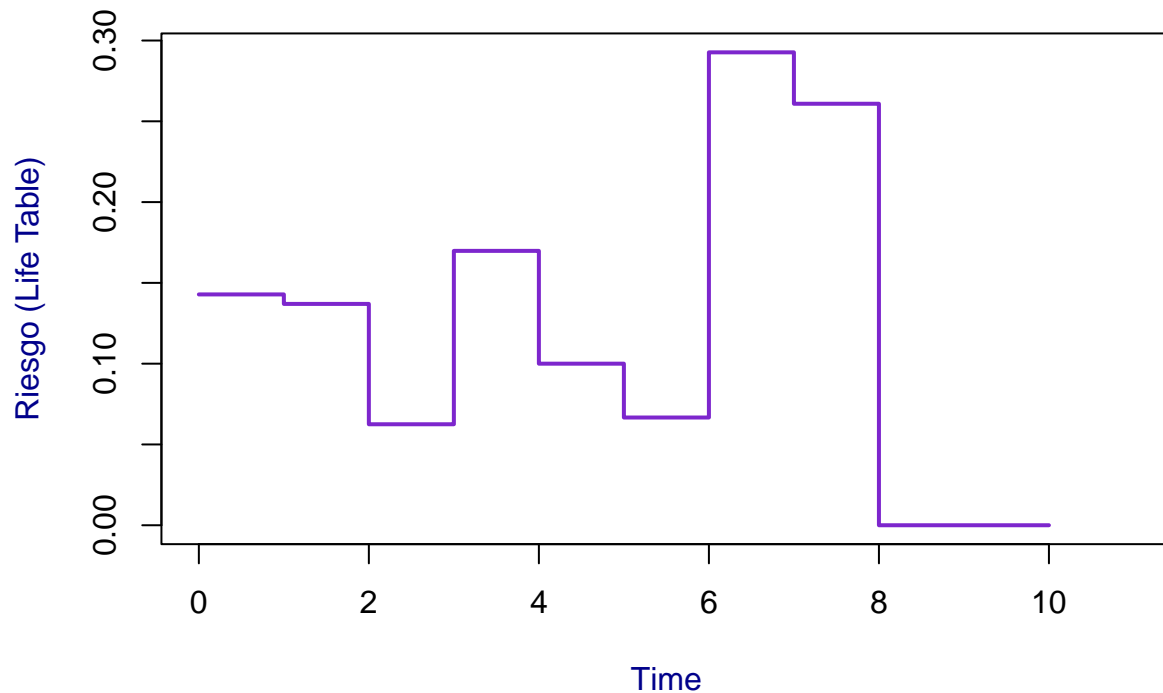


La gráfica de la función de supervivencia acorde a la tabla de vida actuarial parece tener un comportamiento “Casi” lineal. Parece que el mayor salto lo da en el primer mes, y del mes 6 al mes 7. Los menores cambios los presenta, al parecer, del mes 2 al mes 3 y en el periodo de 4 a 6 meses del estudio.

```
### Grafica: funcion de riesgo de la tabla actuarial de vida
```

```
y = rep(Tabla_Vida_Actuarial$hazard, rep(2,11))
plot(x, y, type="l", col="purple3", xlab="Time", ylab="Riesgo (Life Table)",
     col.lab="darkblue",lwd=2)
title(main="Supervivencia al cáncer", cex=.6,col.main="orange")
```


Supervivencia al cáncer



Podemos observar un comportamiento que no es monótono: En los primeros dos meses se mantiene constante, después baja y vuelve a subir constantemente, alcanzando el mayor riesgo entre los 6 y 8 meses, y teniendo un riesgo nulo, de acuerdo a la tabla, a partir del mes 8

Kaplan-Meier

Ejercicio 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.

Creación de la función para K-M

Crearemos una función para facilitar los cálculos:

```
KM<-function(fallas,tiempo,redondear=TRUE,digitos=5){
  long=length(fallas)
  d<-fallas
  f_aux<-rep(1:long)
  s<-rep(1,long)
  var<-rep(1,long)
  f2<-rep(0,long)
  n=seq(long,1,by=-1)
  c=abs(fallas-1)
  for(i in 1:long){
```

```

    if(n[i]>d[i]){
      f2[i]<-d[i]/(n[i]*(n[i]-d[i]))
    }
    else{
      f2[i]=0
    }
  }
  for (i in 1:long){
    f_aux[i]<-(n[i]-d[i])/n[i]
    for(j in 1:i){
      s[i]<-s[i]*f_aux[i]
    }
    for ( j in 1:i){
      var[i]<-sum(f2[1:j])*(s[i])^2
    }
  }
  L<-rep(0,long)
  U<-rep(0,long)
  for (i in 1:long){
    L[i]<-s[i] -qnorm(.975)*sqrt(var[i])
    if(L[i]<0){
      L[i]=0 #Truncamos
    }
    U[i]<-s[i] +qnorm(.975)*sqrt(var[i])
    if(U[i]>1){
      U[i]=1 #Truncamos
    }
  }
  #A tiempo 0 tenemos lo siguiente, por eso añadimos:
  tiempo=c(0,tiempo)
  n=c(long,n)
  d=c(0,d)
  c=c(0,c)
  f_aux=c(1,f_aux)
  s=c(1,s)
  var=c(1,var)
  se=sqrt(var)
  L=c(1,L)
  U=c(1,U)
  KM=data.frame("Tiempo"=tiempo,"nj"=n,"dj"=d,"cj"=c,
                "dif_ponderada"=f_aux,"s(t)"=s,"Var"=var,
                "Se"=se,"L"=L,"U"=U)
  #Agregamos esta opción de redondear para hacerlo más legible
  if(redondear){
    KM=round(KM,digitos)
  }
  return(KM)
}

```

S(t) poblacional

```

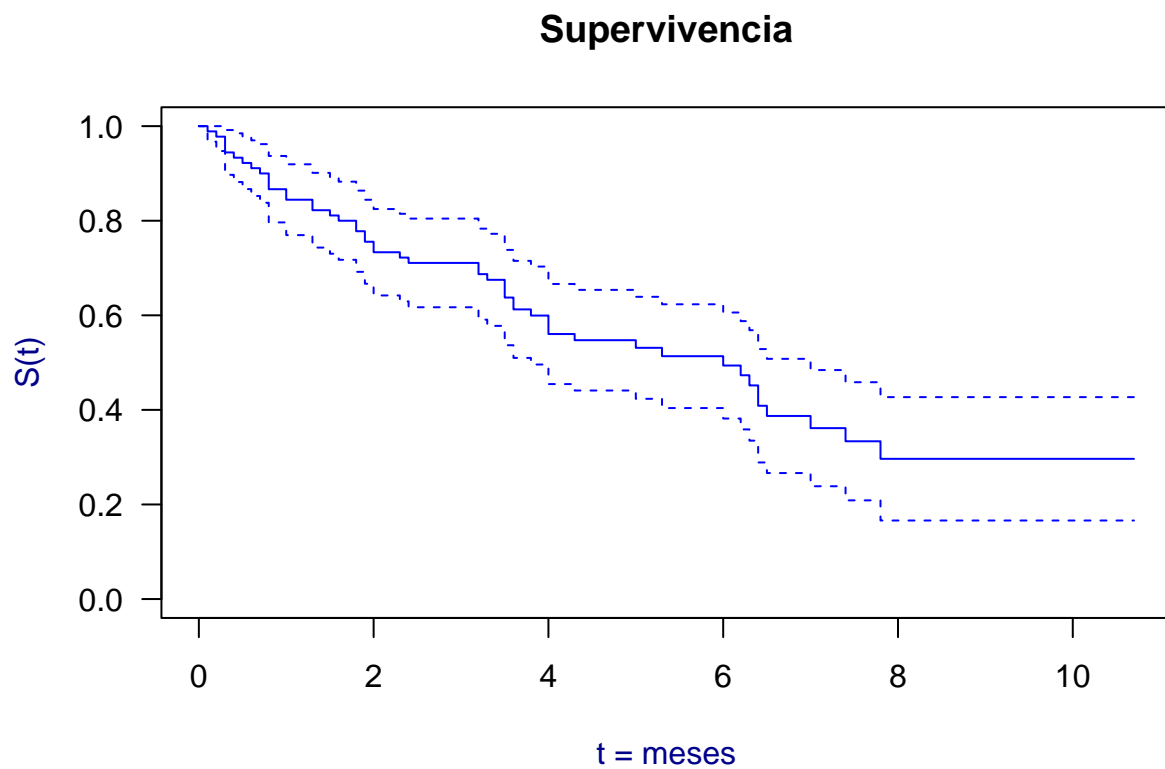
Surv_general <- Surv(larynx$time, larynx$delta,type = "right")
general_fit <- survfit(Surv_general ~ 1, type = "kaplan-meier", conf.int = 0.95,

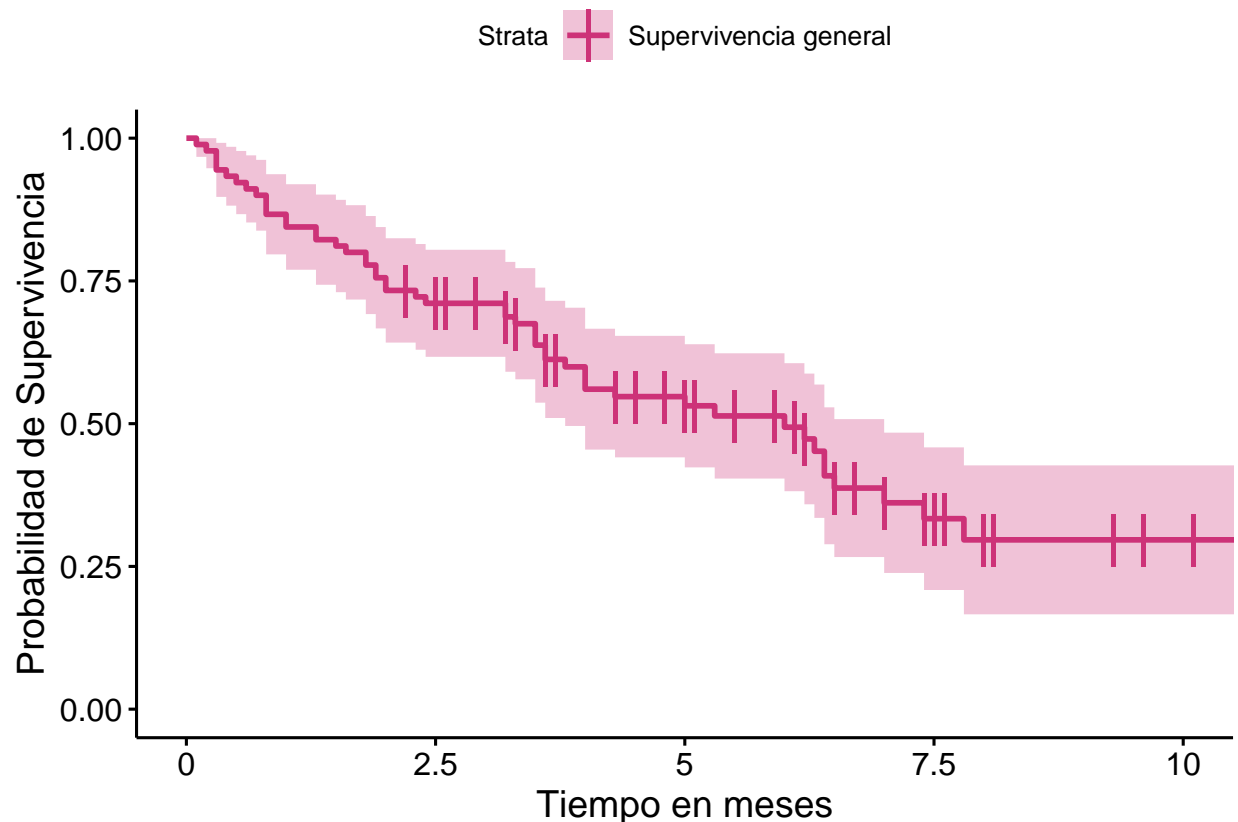
```

```

conf.type = "plain",data=larynx)
plot(general_fit, main="Supervivencia", xlab = "t = meses", ylab = "S(t)",
     col="blue1", las=1, col.lab="darkblue")
ggsurvplot(general_fit, data=larynx,palette = "violetred3",censor.size=7,
           censor.shape= 124,legend.labs = "Supervivencia general",
           xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia",
           col.lab="deeppink")

```





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.

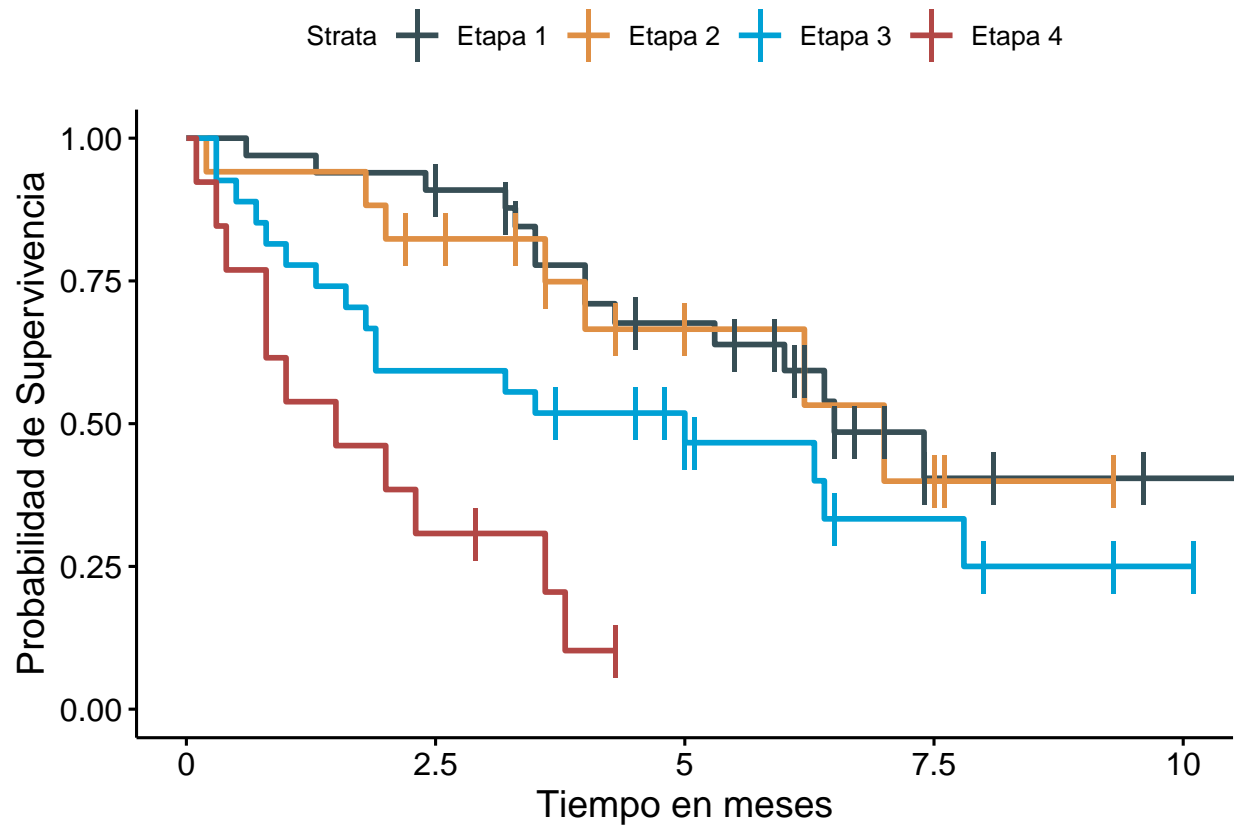
```
KM(delta,time)
```

##	Tiempo	nj	dj	cj	dif_ponderada	s.t.	Var	Se	L	U
## 1	0.0	90	0	0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
## 2	0.6	90	1	0	0.98889	0.98889	0.00012	0.01105	0.96723	1.00000
## 3	1.3	89	1	0	0.98876	0.97765	0.00024	0.01554	0.94720	1.00000
## 4	2.4	88	1	0	0.98864	0.96630	0.00036	0.01891	0.92922	1.00000
## 5	2.5	87	0	1	1.00000	1.00000	0.00038	0.01957	0.96164	1.00000
## 6	3.2	86	1	0	0.98837	0.94320	0.00046	0.02151	0.90104	0.98535
## 7	3.2	85	0	1	1.00000	1.00000	0.00052	0.02280	0.95531	1.00000
## 8	3.3	84	1	0	0.98810	0.91958	0.00056	0.02368	0.87316	0.96601
## 9	3.3	83	0	1	1.00000	1.00000	0.00066	0.02576	0.94952	1.00000
## 10	3.5	82	1	0	0.98780	0.89545	0.00065	0.02555	0.84538	0.94552
## 11	3.5	81	1	0	0.98765	0.88318	0.00076	0.02748	0.82932	0.93704
## 12	4.0	80	1	0	0.98750	0.87078	0.00085	0.02923	0.81350	0.92806
## 13	4.0	79	1	0	0.98734	0.85824	0.00095	0.03081	0.79786	0.91863
## 14	4.3	78	1	0	0.98718	0.84557	0.00104	0.03226	0.78235	0.90879
## 15	4.5	77	0	1	1.00000	1.00000	0.00146	0.03815	0.92523	1.00000
## 16	4.5	76	0	1	1.00000	1.00000	0.00146	0.03815	0.92523	1.00000
## 17	5.3	75	1	0	0.98667	0.80673	0.00106	0.03262	0.74279	0.87067
## 18	5.5	74	0	1	1.00000	1.00000	0.00164	0.04044	0.92074	1.00000
## 19	5.9	73	0	1	1.00000	1.00000	0.00164	0.04044	0.92074	1.00000
## 20	5.9	72	0	1	1.00000	1.00000	0.00164	0.04044	0.92074	1.00000
## 21	6.0	71	1	0	0.98592	0.75300	0.00104	0.03227	0.68975	0.81625
## 22	6.1	70	0	1	1.00000	1.00000	0.00184	0.04286	0.91600	1.00000
## 23	6.2	69	0	1	1.00000	1.00000	0.00184	0.04286	0.91600	1.00000
## 24	6.4	68	1	0	0.98529	0.71124	0.00104	0.03225	0.64803	0.77445
## 25	6.5	67	1	0	0.98507	0.69704	0.00111	0.03330	0.63178	0.76231
## 26	6.5	66	0	1	1.00000	1.00000	0.00228	0.04777	0.90637	1.00000
## 27	6.7	65	0	1	1.00000	1.00000	0.00228	0.04777	0.90637	1.00000
## 28	7.0	64	0	1	1.00000	1.00000	0.00228	0.04777	0.90637	1.00000
## 29	7.4	63	1	0	0.98413	0.63890	0.00104	0.03219	0.57581	0.70199
## 30	7.4	62	0	1	1.00000	1.00000	0.00254	0.05038	0.90125	1.00000
## 31	8.1	61	0	1	1.00000	1.00000	0.00254	0.05038	0.90125	1.00000
## 32	8.1	60	0	1	1.00000	1.00000	0.00254	0.05038	0.90125	1.00000
## 33	9.6	59	0	1	1.00000	1.00000	0.00254	0.05038	0.90125	1.00000
## 34	10.7	58	0	1	1.00000	1.00000	0.00254	0.05038	0.90125	1.00000
## 35	0.2	57	1	0	0.98246	0.54783	0.00086	0.02925	0.49050	0.60517
## 36	1.8	56	1	0	0.98214	0.53225	0.00090	0.03000	0.47345	0.59104
## 37	2.0	55	1	0	0.98182	0.51656	0.00094	0.03062	0.45655	0.57656
## 38	2.2	54	0	1	1.00000	1.00000	0.00351	0.05927	0.88383	1.00000
## 39	2.6	53	0	1	1.00000	1.00000	0.00351	0.05927	0.88383	1.00000
## 40	3.3	52	0	1	1.00000	1.00000	0.00351	0.05927	0.88383	1.00000
## 41	3.6	51	1	0	0.98039	0.45289	0.00080	0.02830	0.39742	0.50836
## 42	3.6	50	0	1	1.00000	1.00000	0.00391	0.06249	0.87752	1.00000
## 43	4.0	49	1	0	0.97959	0.42063	0.00077	0.02768	0.36638	0.47488
## 44	4.3	48	0	1	1.00000	1.00000	0.00433	0.06580	0.87102	1.00000
## 45	4.3	47	0	1	1.00000	1.00000	0.00433	0.06580	0.87102	1.00000
## 46	5.0	46	0	1	1.00000	1.00000	0.00433	0.06580	0.87102	1.00000
## 47	6.2	45	1	0	0.97778	0.35567	0.00061	0.02473	0.30720	0.40414
## 48	7.0	44	1	0	0.97727	0.33942	0.00062	0.02486	0.29070	0.38815
## 49	7.5	43	0	1	1.00000	1.00000	0.00536	0.07324	0.85646	1.00000
## 50	7.6	42	0	1	1.00000	1.00000	0.00536	0.07324	0.85646	1.00000
## 51	9.3	41	0	1	1.00000	1.00000	0.00536	0.07324	0.85646	1.00000
## 52	0.3	40	1	0	0.97500	0.27494	0.00045	0.02131	0.23318	0.31670
## 53	0.3	39	1	0	0.97436	0.25905	0.00045	0.02117	0.21756	0.30055

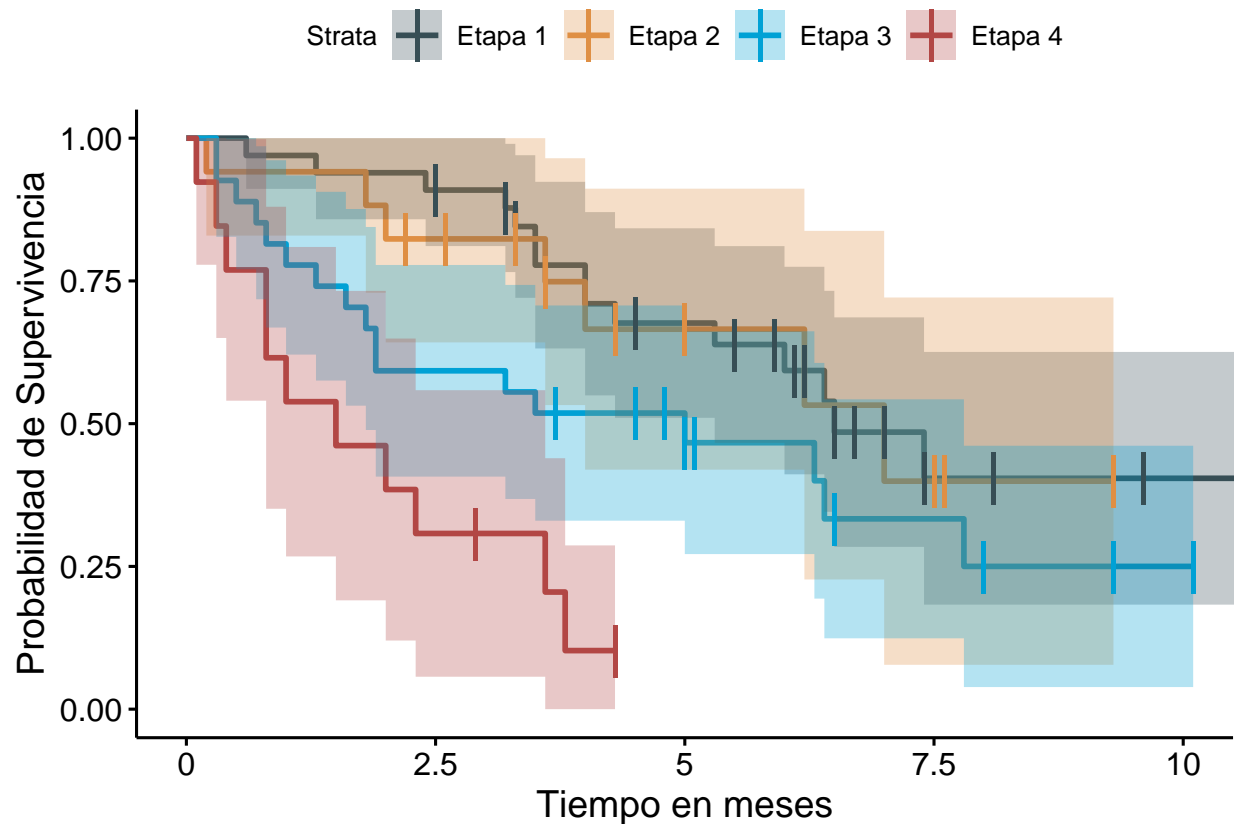
## 54	0.5	38	1	0	0.97368	0.24331	0.00044	0.02092	0.20231	0.28431
## 55	0.7	37	1	0	0.97297	0.22774	0.00042	0.02055	0.18747	0.26802
## 56	0.8	36	1	0	0.97222	0.21238	0.00040	0.02008	0.17303	0.25172
## 57	1.0	35	1	0	0.97143	0.19725	0.00038	0.01950	0.15902	0.23547
## 58	1.3	34	1	0	0.97059	0.18239	0.00035	0.01884	0.14547	0.21931
## 59	1.6	33	1	0	0.96970	0.16784	0.00033	0.01809	0.13239	0.20329
## 60	1.8	32	1	0	0.96875	0.15364	0.00030	0.01726	0.11981	0.18747
## 61	1.9	31	1	0	0.96774	0.13982	0.00027	0.01636	0.10775	0.17189
## 62	1.9	30	1	0	0.96667	0.12644	0.00024	0.01541	0.09624	0.15663
## 63	3.2	29	1	0	0.96552	0.11353	0.00021	0.01440	0.08532	0.14175
## 64	3.5	28	1	0	0.96429	0.10115	0.00018	0.01334	0.07500	0.12730
## 65	3.7	27	0	1	1.00000	1.00000	0.01740	0.13191	0.74146	1.00000
## 66	4.5	26	0	1	1.00000	1.00000	0.01740	0.13191	0.74146	1.00000
## 67	4.8	25	0	1	1.00000	1.00000	0.01740	0.13191	0.74146	1.00000
## 68	4.8	24	0	1	1.00000	1.00000	0.01740	0.13191	0.74146	1.00000
## 69	5.0	23	1	0	0.95652	0.04867	0.00005	0.00677	0.03539	0.06195
## 70	5.0	22	0	1	1.00000	1.00000	0.01938	0.13920	0.72717	1.00000
## 71	5.1	21	0	1	1.00000	1.00000	0.01938	0.13920	0.72717	1.00000
## 72	6.3	20	1	0	0.95000	0.02620	0.00002	0.00389	0.01859	0.03382
## 73	6.4	19	1	0	0.94737	0.02039	0.00001	0.00322	0.01408	0.02670
## 74	6.5	18	0	1	1.00000	1.00000	0.02493	0.15790	0.69052	1.00000
## 75	7.8	17	1	0	0.94118	0.01126	0.00000	0.00191	0.00753	0.01500
## 76	8.0	16	0	1	1.00000	1.00000	0.02861	0.16914	0.66849	1.00000
## 77	9.3	15	0	1	1.00000	1.00000	0.02861	0.16914	0.66849	1.00000
## 78	10.1	14	0	1	1.00000	1.00000	0.02861	0.16914	0.66849	1.00000
## 79	0.1	13	1	0	0.92308	0.00194	0.00000	0.00036	0.00123	0.00266
## 80	0.3	12	1	0	0.91667	0.00103	0.00000	0.00021	0.00062	0.00145
## 81	0.4	11	1	0	0.90909	0.00049	0.00000	0.00011	0.00027	0.00071
## 82	0.8	10	1	0	0.90000	0.00020	0.00000	0.00005	0.00010	0.00029
## 83	0.8	9	1	0	0.88889	0.00006	0.00000	0.00002	0.00003	0.00010
## 84	1.0	8	1	0	0.87500	0.00002	0.00000	0.00000	0.00001	0.00002
## 85	1.5	7	1	0	0.85714	0.00000	0.00000	0.00000	0.00000	0.00000
## 86	2.0	6	1	0	0.83333	0.00000	0.00000	0.00000	0.00000	0.00000
## 87	2.3	5	1	0	0.80000	0.00000	0.00000	0.00000	0.00000	0.00000
## 88	2.9	4	0	1	1.00000	1.00000	0.20169	0.44909	0.11979	1.00000
## 89	3.6	3	1	0	0.66667	0.00000	0.00000	0.00000	0.00000	0.00000
## 90	3.8	2	1	0	0.50000	0.00000	0.00000	0.00000	0.00000	0.00000
## 91	4.3	1	0	1	1.00000	1.00000	0.86835	0.93185	0.00000	1.00000

S(t) por estadio de la enfermedad

```
stage_fit <- survfit(Surv_general ~ stage, type = "kaplan-meier", conf.int = 0.95,
                    conf.type = "plain", data=larynx)
ggsurvplot(stage_fit, data = larynx, palette = pal_jama()(4), censor.size=7,
            censor.shape= 124, legend.labs =c("Etapa 1", "Etapa 2", "Etapa 3",
            "Etapa 4"),
            xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(stage_fit, data = larynx, palette = pal_jama()(4), censor.size=7,
  censor.shape= 124, legend.labs =c("Etapa 1", "Etapa 2", "Etapa 3",
    "Etapa 4"),
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia",
  conf.int = TRUE)
```



Al menos de manera gráfica, vemos que podríamos unir la etapa 1 y 2 en un mismo grupo puesto que parecen describir casi la misma función de supervivencia, y parece haber una diferencia significativa entre la etapa 4, la etapa 3 y, si juntásemos la etapa 1 y 2 en un mismo grupo, estos 3 serían funciones, al menos de manera gráfica, muy distinguibles entre sí. Y hasta cierto punto es intuitivo que aquellos con una etapa más avanzada tengan una menor supervivencia que aquellos en etapas tempranas. Podemos destacar también que los pacientes en etapa 3 parecen tener una alta mortalidad al inicio, al menos en los primeros dos meses, pues observamos fallas constantes.

```
for (i in 1:4){
  print(paste("Para la etapa",i))
  print(KM(delta[stage==i],time[stage==i]))
}
```

```
## [1] "Para la etapa 1"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 33  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 33  1  0      0.96970 0.96970 0.00089 0.02984 0.91121 1.00000
## 3      1.3 32  1  0      0.96875 0.93848 0.00172 0.04150 0.85715 1.00000
## 4      2.4 31  1  0      0.96774 0.90631 0.00249 0.04989 0.80853 1.00000
## 5      2.5 30  0  1      1.00000 1.00000 0.00303 0.05505 0.89211 1.00000
## 6      3.2 29  1  0      0.96552 0.83907 0.00300 0.05478 0.73171 0.94643
## 7      3.2 28  0  1      1.00000 1.00000 0.00426 0.06528 0.87205 1.00000
## 8      3.3 27  1  0      0.96296 0.76783 0.00335 0.05790 0.65435 0.88132
## 9      3.3 26  0  1      1.00000 1.00000 0.00569 0.07541 0.85220 1.00000
## 10     3.5 25  1  0      0.96000 0.69253 0.00353 0.05938 0.57614 0.80893
## 11     3.5 24  1  0      0.95833 0.65338 0.00391 0.06255 0.53079 0.77597
## 12     4.0 23  1  0      0.95652 0.61326 0.00419 0.06473 0.48639 0.74013
```


## 13	4.0	22	1	0	0.95455	0.57222	0.00436	0.06600	0.44285	0.70158
## 14	4.3	21	1	0	0.95238	0.53032	0.00441	0.06642	0.40014	0.66050
## 15	4.5	20	0	1	1.00000	1.00000	0.01569	0.12525	0.75452	1.00000
## 16	4.5	19	0	1	1.00000	1.00000	0.01569	0.12525	0.75452	1.00000
## 17	5.3	18	1	0	0.94444	0.40070	0.00304	0.05517	0.29258	0.50883
## 18	5.5	17	0	1	1.00000	1.00000	0.01895	0.13767	0.73016	1.00000
## 19	5.9	16	0	1	1.00000	1.00000	0.01895	0.13767	0.73016	1.00000
## 20	5.9	15	0	1	1.00000	1.00000	0.01895	0.13767	0.73016	1.00000
## 21	6.0	14	1	0	0.92857	0.22715	0.00126	0.03552	0.15753	0.29676
## 22	6.1	13	0	1	1.00000	1.00000	0.02445	0.15636	0.69354	1.00000
## 23	6.2	12	0	1	1.00000	1.00000	0.02445	0.15636	0.69354	1.00000
## 24	6.4	11	1	0	0.90909	0.11168	0.00042	0.02045	0.07159	0.15176
## 25	6.5	10	1	0	0.90000	0.07977	0.00028	0.01686	0.04673	0.11280
## 26	6.5	9	0	1	1.00000	1.00000	0.04465	0.21131	0.58585	1.00000
## 27	6.7	8	0	1	1.00000	1.00000	0.04465	0.21131	0.58585	1.00000
## 28	7.0	7	0	1	1.00000	1.00000	0.04465	0.21131	0.58585	1.00000
## 29	7.4	6	1	0	0.83333	0.00607	0.00000	0.00169	0.00275	0.00939
## 30	7.4	5	0	1	1.00000	1.00000	0.07798	0.27926	0.45267	1.00000
## 31	8.1	4	0	1	1.00000	1.00000	0.07798	0.27926	0.45267	1.00000
## 32	8.1	3	0	1	1.00000	1.00000	0.07798	0.27926	0.45267	1.00000
## 33	9.6	2	0	1	1.00000	1.00000	0.07798	0.27926	0.45267	1.00000
## 34	10.7	1	0	1	1.00000	1.00000	0.07798	0.27926	0.45267	1.00000

[1] "Para la etapa 2"

##	Tiempo	nj	dj	cj	dif_ponderada	s.t.	Var	Se	L	U
## 1	0.0	17	0	0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
## 2	0.2	17	1	0	0.94118	0.94118	0.00326	0.05707	0.82933	1.00000
## 3	1.8	16	1	0	0.93750	0.87891	0.00606	0.07784	0.72635	1.00000
## 4	2.0	15	1	0	0.93333	0.81304	0.00833	0.09128	0.63413	0.99195
## 5	2.2	14	0	1	1.00000	1.00000	0.01261	0.11227	0.77995	1.00000
## 6	2.6	13	0	1	1.00000	1.00000	0.01261	0.11227	0.77995	1.00000
## 7	3.3	12	0	1	1.00000	1.00000	0.01261	0.11227	0.77995	1.00000
## 8	3.6	11	1	0	0.90909	0.51316	0.00571	0.07559	0.36501	0.66130
## 9	3.6	10	0	1	1.00000	1.00000	0.02170	0.14730	0.71131	1.00000
## 10	4.0	9	1	0	0.88889	0.34644	0.00427	0.06535	0.21835	0.47453
## 11	4.3	8	0	1	1.00000	1.00000	0.03558	0.18864	0.63027	1.00000
## 12	4.3	7	0	1	1.00000	1.00000	0.03558	0.18864	0.63027	1.00000
## 13	5.0	6	0	1	1.00000	1.00000	0.03558	0.18864	0.63027	1.00000
## 14	6.2	5	1	0	0.80000	0.05498	0.00026	0.01608	0.02345	0.08650
## 15	7.0	4	1	0	0.75000	0.01782	0.00005	0.00732	0.00346	0.03217
## 16	7.5	3	0	1	1.00000	1.00000	0.16892	0.41100	0.19446	1.00000
## 17	7.6	2	0	1	1.00000	1.00000	0.16892	0.41100	0.19446	1.00000
## 18	9.3	1	0	1	1.00000	1.00000	0.16892	0.41100	0.19446	1.00000

[1] "Para la etapa 3"

##	Tiempo	nj	dj	cj	dif_ponderada	s.t.	Var	Se	L	U
## 1	0.0	27	0	0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
## 2	0.3	27	1	0	0.96296	0.96296	0.00132	0.03634	0.89173	1.00000
## 3	0.3	26	1	0	0.96154	0.92456	0.00253	0.05033	0.82592	1.00000
## 4	0.5	25	1	0	0.96000	0.88474	0.00362	0.06020	0.76675	1.00000
## 5	0.7	24	1	0	0.95833	0.84346	0.00458	0.06769	0.71079	0.97614
## 6	0.8	23	1	0	0.95652	0.80071	0.00540	0.07346	0.65672	0.94469
## 7	1.0	22	1	0	0.95455	0.75645	0.00606	0.07782	0.60393	0.90896
## 8	1.3	21	1	0	0.95238	0.71068	0.00655	0.08091	0.55209	0.86927
## 9	1.6	20	1	0	0.95000	0.66342	0.00686	0.08285	0.50104	0.82580
## 10	1.8	19	1	0	0.94737	0.61471	0.00700	0.08365	0.45076	0.77866

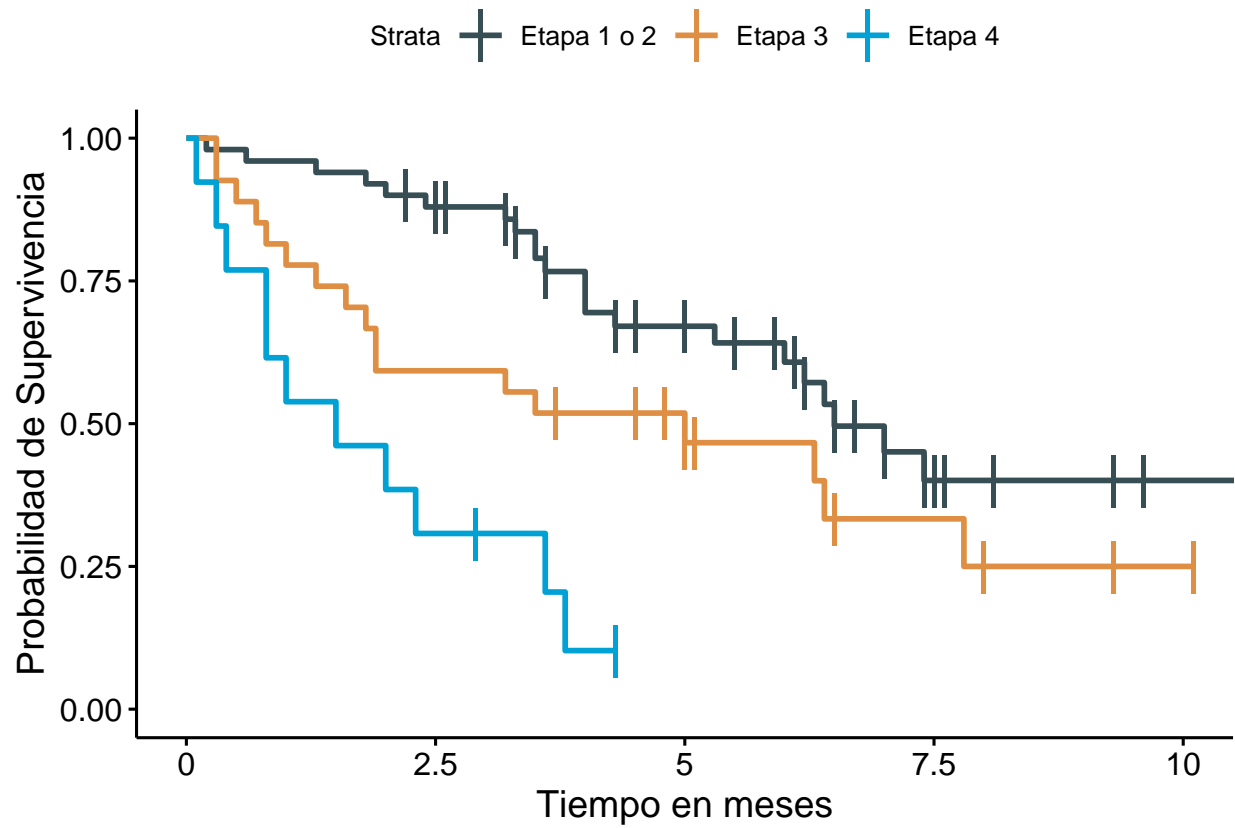
```
## 11    1.9 18  1  0      0.94444 0.56463 0.00695 0.08334 0.40129 0.72798
## 12    1.9 17  1  0      0.94118 0.51331 0.00671 0.08191 0.35277 0.67385
## 13    3.2 16  1  0      0.93750 0.46095 0.00630 0.07934 0.30544 0.61646
## 14    3.5 15  1  0      0.93333 0.40783 0.00572 0.07563 0.25959 0.55606
## 15    3.7 14  0  1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 16    4.5 13  0  1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 17    4.8 12  0  1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 18    4.8 11  0  1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 19    5.0 10  1  0      0.90000 0.15009 0.00103 0.03202 0.08734 0.21285
## 20    5.0  9  0  1      1.00000 1.00000 0.04550 0.21331 0.58191 1.00000
## 21    5.1  8  0  1      1.00000 1.00000 0.04550 0.21331 0.58191 1.00000
## 22    6.3  7  1  0      0.85714 0.03928 0.00011 0.01034 0.01901 0.05954
## 23    6.4  6  1  0      0.83333 0.01811 0.00003 0.00580 0.00674 0.02949
## 24    6.5  5  0  1      1.00000 1.00000 0.10265 0.32038 0.37206 1.00000
## 25    7.8  4  1  0      0.75000 0.00100 0.00000 0.00043 0.00016 0.00185
## 26    8.0  3  0  1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
## 27    9.3  2  0  1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
## 28   10.1  1  0  1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
```

```
## [1] "Para la etapa 4"
```

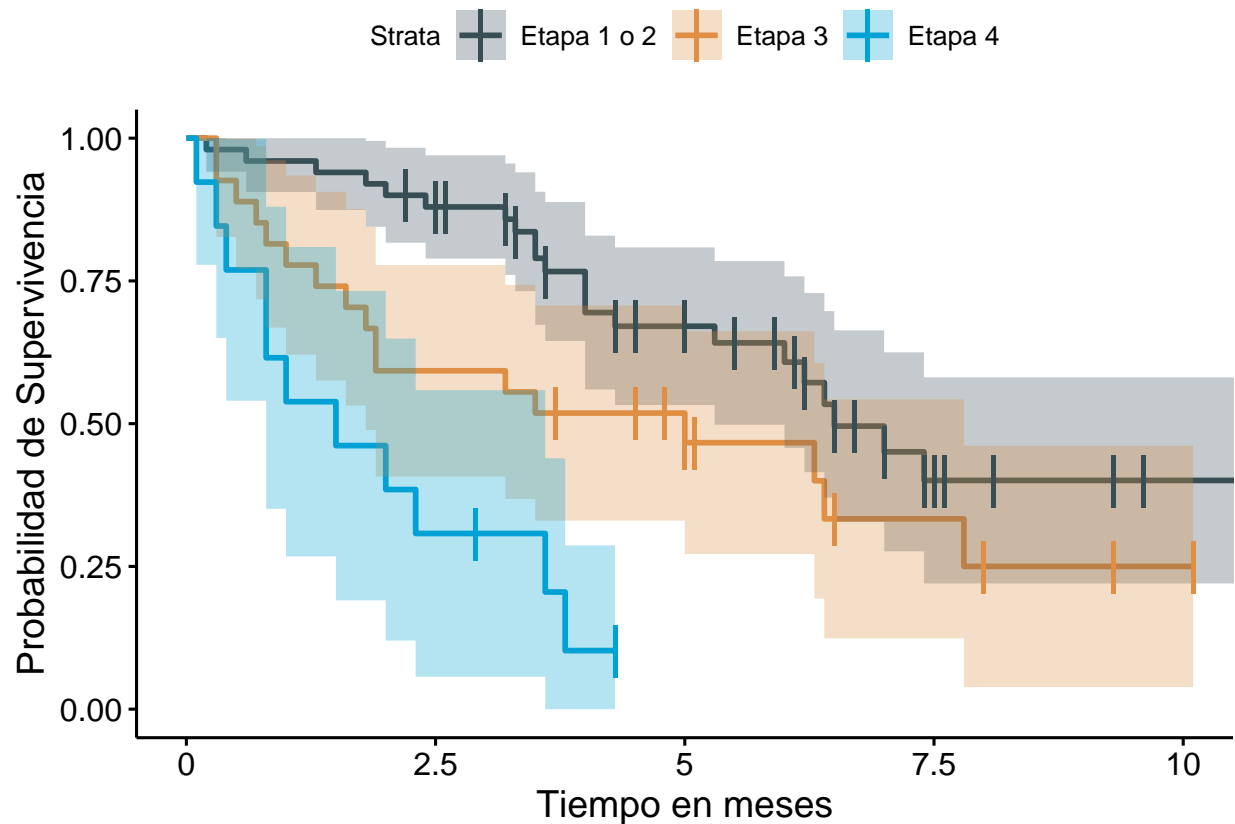
```
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 13  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.1 13  1  0      0.92308 0.92308 0.00546 0.07391 0.77823 1.00000
## 3      0.3 12  1  0      0.91667 0.84028 0.00988 0.09937 0.64551 1.00000
## 4      0.4 11  1  0      0.90909 0.75131 0.01303 0.11413 0.52762 0.97501
## 5      0.8 10  1  0      0.90000 0.65610 0.01472 0.12131 0.41833 0.89387
## 6      0.8  9  1  0      0.88889 0.55493 0.01481 0.12168 0.31645 0.79341
## 7      1.0  8  1  0      0.87500 0.44880 0.01328 0.11524 0.22293 0.67466
## 8      1.5  7  1  0      0.85714 0.33992 0.01037 0.10183 0.14033 0.53950
## 9      2.0  6  1  0      0.83333 0.23257 0.00666 0.08159 0.07265 0.39248
## 10     2.3  5  1  0      0.80000 0.13422 0.00312 0.05584 0.02478 0.24366
## 11     2.9  4  0  1      1.00000 1.00000 0.17308 0.41603 0.18461 1.00000
## 12     3.6  3  1  0      0.66667 0.01156 0.00005 0.00674 0.00000 0.02477
## 13     3.8  2  1  0      0.50000 0.00024 0.00000 0.00022 0.00000 0.00068
## 14     4.3  1  0  1      1.00000 1.00000 0.83974 0.91638 0.00000 1.00000
```

¿Y si agrupamos la etapa 1 y 2? Ya que, de manera gráfica vimos que se parecían mucho sus funciones

```
stage_aux<- larynx %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"1 o 2",stage)))
stage_fit_aux <- survfit(Surv_general ~ stage_aux, type = "kaplan-meier", conf.int = 0.95,
  conf.type = "plain",data=stage_aux)
ggsurvplot(stage_fit_aux,data = stage_aux,palette = pal_jama()(4),censor.size=7,
  censor.shape= 124,legend.labs =c("Etapa 1 o 2", "Etapa 3", "Etapa 4"),
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(stage_fit_aux, data = stage_aux, palette = pal_jama()(4), censor.size=7,
  censor.shape= 124, legend.labs =c("Etapa 1 o 2", "Etapa 3", "Etapa 4"),
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia",
  conf.int = TRUE)
```



```
stage_aux %>%
  group_by(stage_aux) %>%
  summarise(n=n(),perc=n/0.9)
```

```
## # A tibble: 3 x 3
##   stage_aux     n perc
##   <fct>       <int> <dbl>
## 1 1 o 2         50  55.6
## 2 3           27  30
## 3 4           13  14.4
```

¡Sí se ven bastante diferentes! Aunque, evidentemente, se tiene una mayor concentración de datos en el primer grupo, que es la etapa 1 y 2 juntas, a comparación, por ejemplo, de la etapa 4: Poco más de la mitad se encuentran en la etapa 1 o 2

```
for(i in levels(stage_aux$stage_aux)){
  print(paste("Para la etapa:",i))
  delta_aux=(stage_aux$delta)[stage_aux$stage_aux==i]
  time_aux=(stage_aux$time)[stage_aux$stage_aux==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "Para la etapa: 1 o 2"
##   Tiempo nj dj cj dif_ponderada   s.t.   Var   Se   L   U
## 1   0.0 50  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2   0.6 50  1  0      0.98000 0.98000 0.00039 0.01980 0.94119 1.00000
## 3   1.3 49  1  0      0.97959 0.95960 0.00077 0.02770 0.90531 1.00000
## 4   2.4 48  1  0      0.97917 0.93879 0.00113 0.03354 0.87305 1.00000
```

## 5	2.5	47	0	1	1.00000	1.00000	0.00128	0.03573	0.92997	1.00000
## 6	3.2	46	1	0	0.97826	0.89593	0.00141	0.03758	0.82227	0.96959
## 7	3.2	45	0	1	1.00000	1.00000	0.00176	0.04195	0.91778	1.00000
## 8	3.3	44	1	0	0.97727	0.85135	0.00166	0.04072	0.77154	0.93117
## 9	3.3	43	0	1	1.00000	1.00000	0.00229	0.04784	0.90624	1.00000
## 10	3.5	42	1	0	0.97619	0.80503	0.00186	0.04312	0.72052	0.88954
## 11	3.5	41	1	0	0.97561	0.78120	0.00212	0.04608	0.69089	0.87150
## 12	4.0	40	1	0	0.97500	0.75692	0.00236	0.04858	0.66170	0.85214
## 13	4.0	39	1	0	0.97436	0.73220	0.00257	0.05070	0.63283	0.83157
## 14	4.3	38	1	0	0.97368	0.70703	0.00275	0.05246	0.60420	0.80985
## 15	4.5	37	0	1	1.00000	1.00000	0.00551	0.07420	0.85457	1.00000
## 16	4.5	36	0	1	1.00000	1.00000	0.00551	0.07420	0.85457	1.00000
## 17	5.3	35	1	0	0.97143	0.62889	0.00251	0.05010	0.53070	0.72708
## 18	5.5	34	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 19	5.9	33	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 20	5.9	32	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 21	6.0	31	1	0	0.96774	0.51903	0.00200	0.04471	0.43139	0.60666
## 22	6.1	30	0	1	1.00000	1.00000	0.00742	0.08615	0.83115	1.00000
## 23	6.2	29	0	1	1.00000	1.00000	0.00742	0.08615	0.83115	1.00000
## 24	6.4	28	1	0	0.96429	0.43324	0.00164	0.04051	0.35384	0.51265
## 25	6.5	27	1	0	0.96296	0.40423	0.00166	0.04076	0.32434	0.48412
## 26	6.5	26	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 27	6.7	25	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 28	7.0	24	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 29	7.4	23	1	0	0.95652	0.28804	0.00101	0.03174	0.22583	0.35026
## 30	7.4	22	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 31	8.1	21	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 32	8.1	20	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 33	9.6	19	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 34	10.7	18	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 35	0.2	17	1	0	0.94118	0.12730	0.00026	0.01601	0.09591	0.15868
## 36	1.8	16	1	0	0.93750	0.10447	0.00022	0.01477	0.07552	0.13342
## 37	2.0	15	1	0	0.93333	0.08343	0.00017	0.01313	0.05771	0.10916
## 38	2.2	14	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 39	2.6	13	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 40	3.3	12	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 41	3.6	11	1	0	0.90909	0.02209	0.00002	0.00406	0.01413	0.03006
## 42	3.6	10	0	1	1.00000	1.00000	0.03384	0.18396	0.63945	1.00000
## 43	4.0	9	1	0	0.88889	0.00711	0.00000	0.00155	0.00406	0.01015
## 44	4.3	8	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 45	4.3	7	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 46	5.0	6	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 47	6.2	5	1	0	0.80000	0.00003	0.00000	0.00001	0.00001	0.00006
## 48	7.0	4	1	0	0.75000	0.00000	0.00000	0.00000	0.00000	0.00000
## 49	7.5	3	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000
## 50	7.6	2	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000
## 51	9.3	1	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000
## [1] "Para la etapa: 3"										
##	Tiempo	nj	dj	cj	dif_ponderada	s.t.	Var	Se	L	U
## 1	0.0	27	0	0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
## 2	0.3	27	1	0	0.96296	0.96296	0.00132	0.03634	0.89173	1.00000
## 3	0.3	26	1	0	0.96154	0.92456	0.00253	0.05033	0.82592	1.00000
## 4	0.5	25	1	0	0.96000	0.88474	0.00362	0.06020	0.76675	1.00000
## 5	0.7	24	1	0	0.95833	0.84346	0.00458	0.06769	0.71079	0.97614

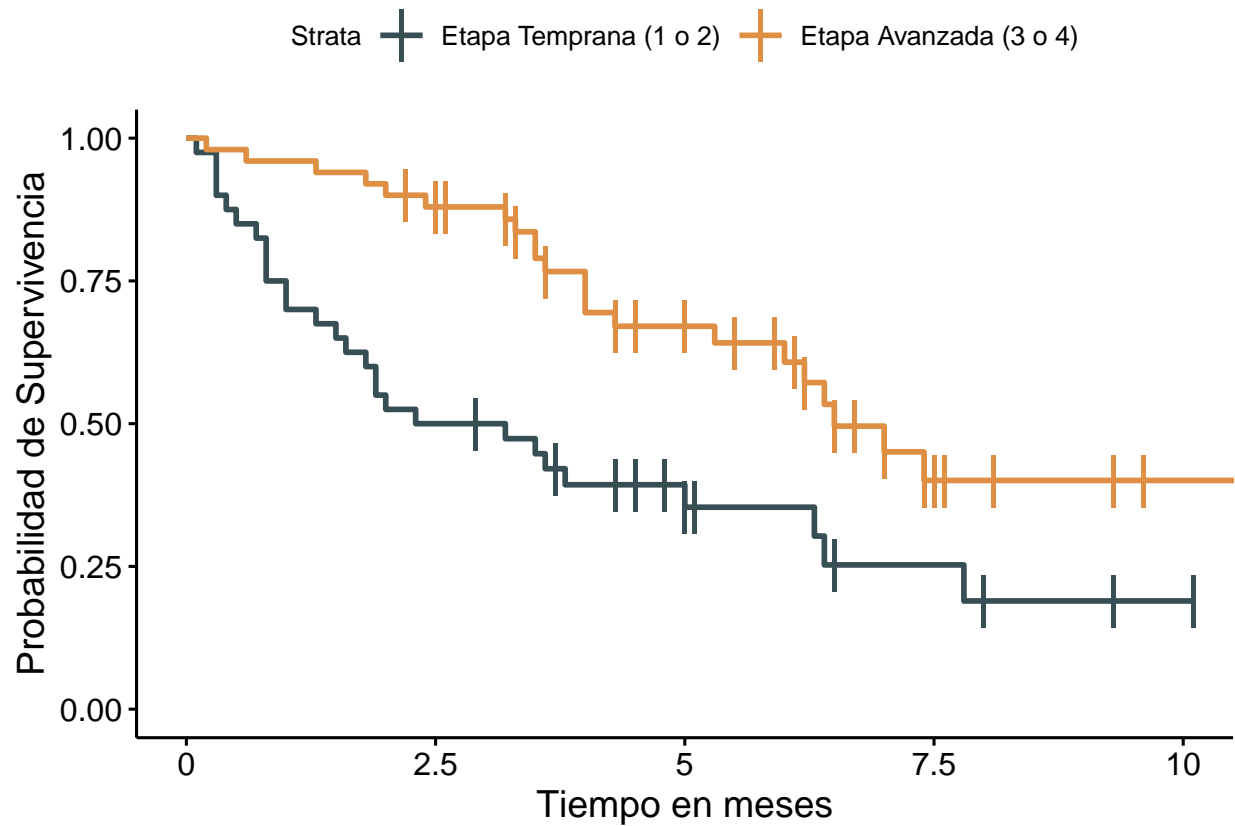
```
## 6      0.8 23 1 0      0.95652 0.80071 0.00540 0.07346 0.65672 0.94469
## 7      1.0 22 1 0      0.95455 0.75645 0.00606 0.07782 0.60393 0.90896
## 8      1.3 21 1 0      0.95238 0.71068 0.00655 0.08091 0.55209 0.86927
## 9      1.6 20 1 0      0.95000 0.66342 0.00686 0.08285 0.50104 0.82580
## 10     1.8 19 1 0      0.94737 0.61471 0.00700 0.08365 0.45076 0.77866
## 11     1.9 18 1 0      0.94444 0.56463 0.00695 0.08334 0.40129 0.72798
## 12     1.9 17 1 0      0.94118 0.51331 0.00671 0.08191 0.35277 0.67385
## 13     3.2 16 1 0      0.93750 0.46095 0.00630 0.07934 0.30544 0.61646
## 14     3.5 15 1 0      0.93333 0.40783 0.00572 0.07563 0.25959 0.55606
## 15     3.7 14 0 1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 16     4.5 13 0 1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 17     4.8 12 0 1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 18     4.8 11 0 1      1.00000 1.00000 0.03439 0.18545 0.63653 1.00000
## 19     5.0 10 1 0      0.90000 0.15009 0.00103 0.03202 0.08734 0.21285
## 20     5.0 9 0 1      1.00000 1.00000 0.04550 0.21331 0.58191 1.00000
## 21     5.1 8 0 1      1.00000 1.00000 0.04550 0.21331 0.58191 1.00000
## 22     6.3 7 1 0      0.85714 0.03928 0.00011 0.01034 0.01901 0.05954
## 23     6.4 6 1 0      0.83333 0.01811 0.00003 0.00580 0.00674 0.02949
## 24     6.5 5 0 1      1.00000 1.00000 0.10265 0.32038 0.37206 1.00000
## 25     7.8 4 1 0      0.75000 0.00100 0.00000 0.00043 0.00016 0.00185
## 26     8.0 3 0 1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
## 27     9.3 2 0 1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
## 28    10.1 1 0 1      1.00000 1.00000 0.18598 0.43125 0.15476 1.00000
```

```
## [1] "Para la etapa: 4"
```

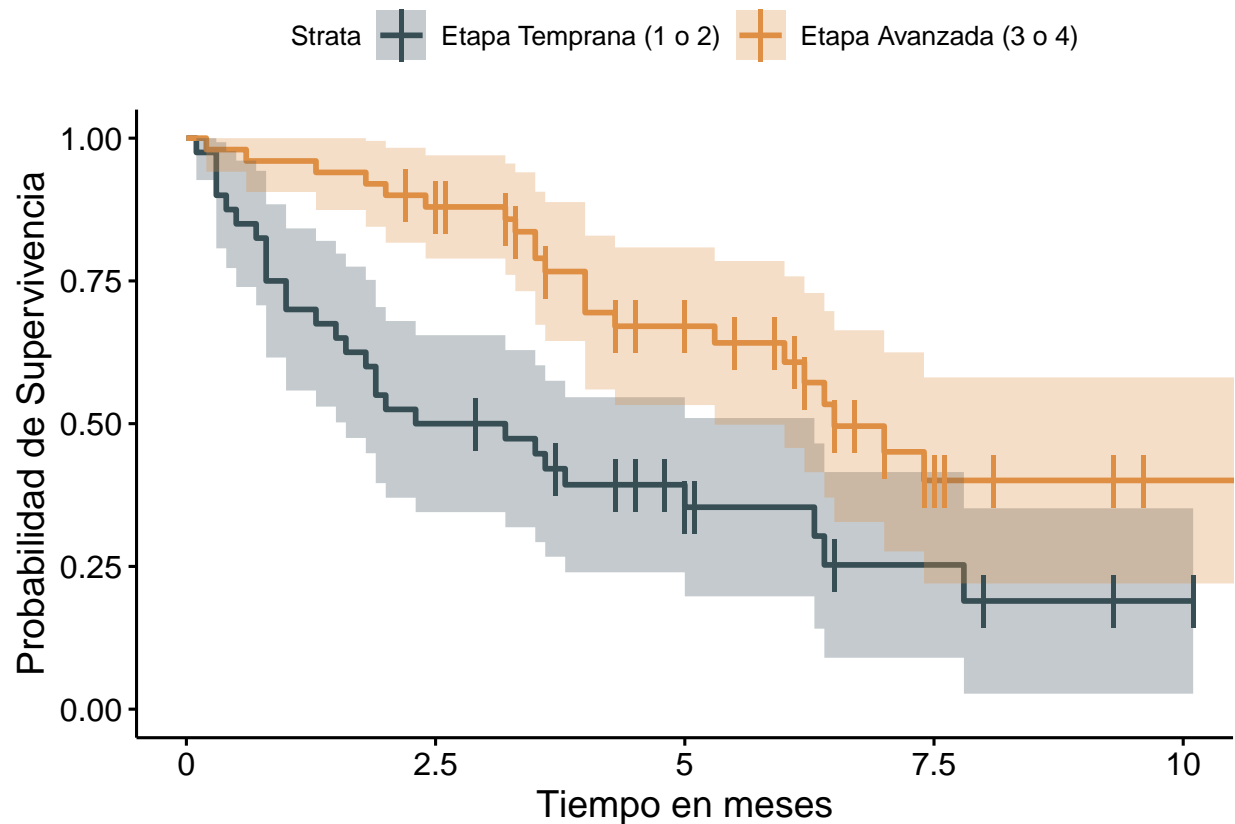
```
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 13 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.1 13 1 0      0.92308 0.92308 0.00546 0.07391 0.77823 1.00000
## 3      0.3 12 1 0      0.91667 0.84028 0.00988 0.09937 0.64551 1.00000
## 4      0.4 11 1 0      0.90909 0.75131 0.01303 0.11413 0.52762 0.97501
## 5      0.8 10 1 0      0.90000 0.65610 0.01472 0.12131 0.41833 0.89387
## 6      0.8 9 1 0      0.88889 0.55493 0.01481 0.12168 0.31645 0.79341
## 7      1.0 8 1 0      0.87500 0.44880 0.01328 0.11524 0.22293 0.67466
## 8      1.5 7 1 0      0.85714 0.33992 0.01037 0.10183 0.14033 0.53950
## 9      2.0 6 1 0      0.83333 0.23257 0.00666 0.08159 0.07265 0.39248
## 10     2.3 5 1 0      0.80000 0.13422 0.00312 0.05584 0.02478 0.24366
## 11     2.9 4 0 1      1.00000 1.00000 0.17308 0.41603 0.18461 1.00000
## 12     3.6 3 1 0      0.66667 0.01156 0.00005 0.00674 0.00000 0.02477
## 13     3.8 2 1 0      0.50000 0.00024 0.00000 0.00022 0.00000 0.00068
## 14     4.3 1 0 1      1.00000 1.00000 0.83974 0.91638 0.00000 1.00000
```

Tratemos con una tercera clasificación, por curiosidad, de qué pasaría si clasificamos a la 1 y 2 como “Etapas tempranas” Y la 3 y 4 como “Etapas avanzadas”, ya que además así quedaría casi la misma proporción de datos:

```
stage_aux2<- larynx %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"Temprana(1,2)",
                                "Avanzada(3,4)")))
stage_fit_aux <- survfit(Surv_general ~ stage_aux, type = "kaplan-meier", conf.int = 0.95,
                        conf.type = "plain",data=stage_aux2)
ggsurvplot(stage_fit_aux,data = stage_aux2,palette = pal_jama()(4),censor.size=7,
            censor.shape= 124,legend.labs =c("Etapa Temprana (1 o 2)",
                                              "Etapa Avanzada (3 o 4)"),
            xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(stage_fit_aux, data = stage_aux2, palette = pal_jama()(4), censor.size=7,
  censor.shape= 124, legend.labs =c("Etapa Temprana (1 o 2)",
                                     "Etapa Avanzada (3 o 4)"),
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia",
  conf.int = TRUE)
```



```
stage_aux2 %>%
  group_by(stage_aux) %>%
  summarise(n=n(),perc=n/0.9)
```

```
## # A tibble: 2 x 3
##   stage_aux      n perc
##   <fct>      <int> <dbl>
## 1 Avanzada(3,4)    40  44.4
## 2 Temprana(1,2)   50  55.6
```

Bajo esta clasificación los intervalos de confianza casi ni se tocan y tenemos casi la misma proporción en cada clasificación

```
for(i in levels(stage_aux2$stage_aux)){
  print(paste("Para la etapa:",i))
  delta_aux=(stage_aux2$delta)[stage_aux2$stage_aux==i]
  time_aux=(stage_aux2$time)[stage_aux2$stage_aux==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "Para la etapa: Avanzada(3,4)"
##   Tiempo nj dj cj dif_ponderada s.t. Var Se L U
## 1 0.0 40 0 0 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2 0.3 40 1 0 0.97500 0.97500 0.00061 0.02469 0.92662 1.00000
## 3 0.3 39 1 0 0.97436 0.94938 0.00119 0.03444 0.88188 1.00000
## 4 0.5 38 1 0 0.97368 0.92311 0.00173 0.04156 0.84165 1.00000
## 5 0.7 37 1 0 0.97297 0.89620 0.00223 0.04723 0.80362 0.98877
## 6 0.8 36 1 0 0.97222 0.86862 0.00269 0.05191 0.76687 0.97036
```



```

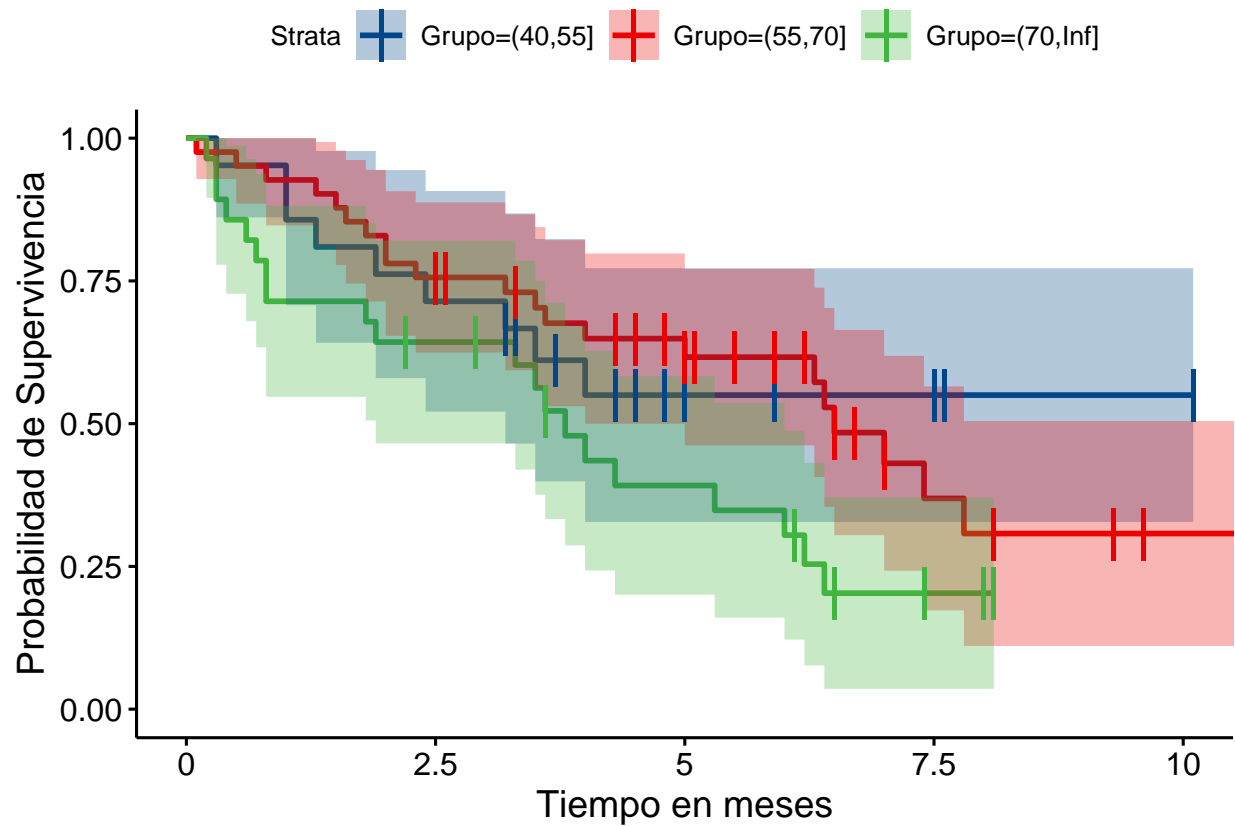
## 7      1.0 35 1 0      0.97143 0.84036 0.00312 0.05582 0.73096 0.94976
## 8      1.3 34 1 0      0.97059 0.81142 0.00349 0.05909 0.69561 0.92723
## 9      1.6 33 1 0      0.96970 0.78179 0.00382 0.06181 0.66065 0.90292
## 10     1.8 32 1 0      0.96875 0.75146 0.00410 0.06402 0.62598 0.87694
## 11     1.9 31 1 0      0.96774 0.72044 0.00433 0.06577 0.59154 0.84934
## 12     1.9 30 1 0      0.96667 0.68872 0.00450 0.06707 0.55727 0.82017
## 13     3.2 29 1 0      0.96552 0.65633 0.00462 0.06794 0.52317 0.78948
## 14     3.5 28 1 0      0.96429 0.62327 0.00468 0.06838 0.48924 0.75729
## 15     3.7 27 0 1      1.00000 1.00000 0.01204 0.10971 0.78497 1.00000
## 16     4.5 26 0 1      1.00000 1.00000 0.01204 0.10971 0.78497 1.00000
## 17     4.8 25 0 1      1.00000 1.00000 0.01204 0.10971 0.78497 1.00000
## 18     4.8 24 0 1      1.00000 1.00000 0.01204 0.10971 0.78497 1.00000
## 19     5.0 23 1 0      0.95652 0.44927 0.00283 0.05318 0.34503 0.55351
## 20     5.0 22 0 1      1.00000 1.00000 0.01401 0.11838 0.76798 1.00000
## 21     5.1 21 0 1      1.00000 1.00000 0.01401 0.11838 0.76798 1.00000
## 22     6.3 20 1 0      0.95000 0.34056 0.00193 0.04394 0.25445 0.42668
## 23     6.4 19 1 0      0.94737 0.30438 0.00181 0.04258 0.22093 0.38783
## 24     6.5 18 0 1      1.00000 1.00000 0.01957 0.13989 0.72582 1.00000
## 25     7.8 17 1 0      0.94118 0.23340 0.00127 0.03559 0.16366 0.30315
## 26     8.0 16 0 1      1.00000 1.00000 0.02325 0.15246 0.70118 1.00000
## 27     9.3 15 0 1      1.00000 1.00000 0.02325 0.15246 0.70118 1.00000
## 28    10.1 14 0 1      1.00000 1.00000 0.02325 0.15246 0.70118 1.00000
## 29     0.1 13 1 0      0.92308 0.10633 0.00034 0.01831 0.07044 0.14222
## 30     0.3 12 1 0      0.91667 0.08019 0.00024 0.01547 0.04986 0.11052
## 31     0.4 11 1 0      0.90909 0.05731 0.00015 0.01233 0.03313 0.08148
## 32     0.8 10 1 0      0.90000 0.03815 0.00008 0.00914 0.02023 0.05607
## 33     0.8 9 1 0      0.88889 0.02307 0.00004 0.00616 0.01100 0.03515
## 34     1.0 8 1 0      0.87500 0.01220 0.00001 0.00364 0.00506 0.01934
## 35     1.5 7 1 0      0.85714 0.00529 0.00000 0.00178 0.00181 0.00878
## 36     2.0 6 1 0      0.83333 0.00169 0.00000 0.00065 0.00042 0.00296
## 37     2.3 5 1 0      0.80000 0.00032 0.00000 0.00014 0.00004 0.00061
## 38     2.9 4 0 1      1.00000 1.00000 0.19632 0.44308 0.13157 1.00000
## 39     3.6 3 1 0      0.66667 0.00000 0.00000 0.00000 0.00000 0.00000
## 40     3.8 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 41     4.3 1 0 1      1.00000 1.00000 0.86299 0.92897 0.00000 1.00000
## [1] "Para la etapa: Temprana(1,2)"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 50 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 50 1 0      0.98000 0.98000 0.00039 0.01980 0.94119 1.00000
## 3      1.3 49 1 0      0.97959 0.95960 0.00077 0.02770 0.90531 1.00000
## 4      2.4 48 1 0      0.97917 0.93879 0.00113 0.03354 0.87305 1.00000
## 5      2.5 47 0 1      1.00000 1.00000 0.00128 0.03573 0.92997 1.00000
## 6      3.2 46 1 0      0.97826 0.89593 0.00141 0.03758 0.82227 0.96959
## 7      3.2 45 0 1      1.00000 1.00000 0.00176 0.04195 0.91778 1.00000
## 8      3.3 44 1 0      0.97727 0.85135 0.00166 0.04072 0.77154 0.93117
## 9      3.3 43 0 1      1.00000 1.00000 0.00229 0.04784 0.90624 1.00000
## 10     3.5 42 1 0      0.97619 0.80503 0.00186 0.04312 0.72052 0.88954
## 11     3.5 41 1 0      0.97561 0.78120 0.00212 0.04608 0.69089 0.87150
## 12     4.0 40 1 0      0.97500 0.75692 0.00236 0.04858 0.66170 0.85214
## 13     4.0 39 1 0      0.97436 0.73220 0.00257 0.05070 0.63283 0.83157
## 14     4.3 38 1 0      0.97368 0.70703 0.00275 0.05246 0.60420 0.80985
## 15     4.5 37 0 1      1.00000 1.00000 0.00551 0.07420 0.85457 1.00000
## 16     4.5 36 0 1      1.00000 1.00000 0.00551 0.07420 0.85457 1.00000
## 17     5.3 35 1 0      0.97143 0.62889 0.00251 0.05010 0.53070 0.72708

```

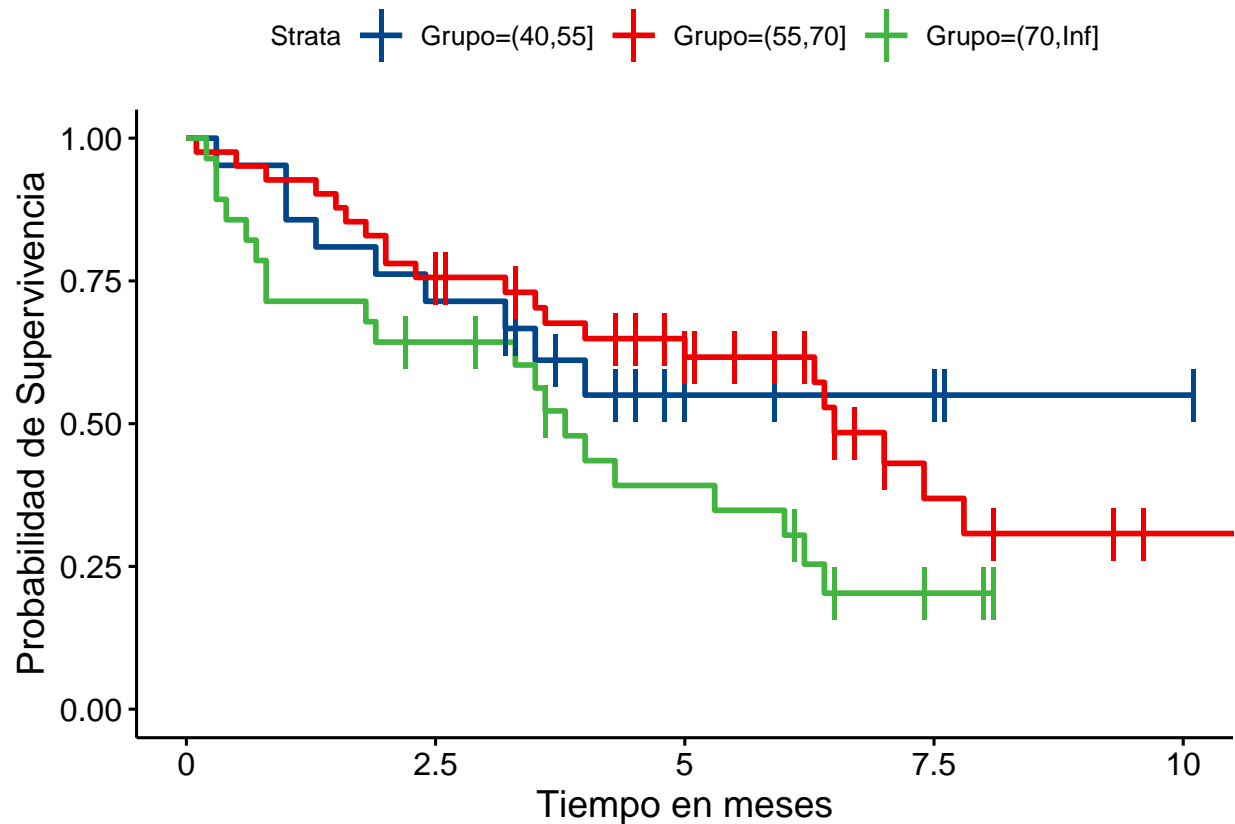
## 18	5.5	34	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 19	5.9	33	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 20	5.9	32	0	1	1.00000	1.00000	0.00635	0.07966	0.84386	1.00000
## 21	6.0	31	1	0	0.96774	0.51903	0.00200	0.04471	0.43139	0.60666
## 22	6.1	30	0	1	1.00000	1.00000	0.00742	0.08615	0.83115	1.00000
## 23	6.2	29	0	1	1.00000	1.00000	0.00742	0.08615	0.83115	1.00000
## 24	6.4	28	1	0	0.96429	0.43324	0.00164	0.04051	0.35384	0.51265
## 25	6.5	27	1	0	0.96296	0.40423	0.00166	0.04076	0.32434	0.48412
## 26	6.5	26	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 27	6.7	25	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 28	7.0	24	0	1	1.00000	1.00000	0.01017	0.10084	0.80236	1.00000
## 29	7.4	23	1	0	0.95652	0.28804	0.00101	0.03174	0.22583	0.35026
## 30	7.4	22	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 31	8.1	21	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 32	8.1	20	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 33	9.6	19	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 34	10.7	18	0	1	1.00000	1.00000	0.01214	0.11020	0.78400	1.00000
## 35	0.2	17	1	0	0.94118	0.12730	0.00026	0.01601	0.09591	0.15868
## 36	1.8	16	1	0	0.93750	0.10447	0.00022	0.01477	0.07552	0.13342
## 37	2.0	15	1	0	0.93333	0.08343	0.00017	0.01313	0.05771	0.10916
## 38	2.2	14	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 39	2.6	13	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 40	3.3	12	0	1	1.00000	1.00000	0.02475	0.15732	0.69166	1.00000
## 41	3.6	11	1	0	0.90909	0.02209	0.00002	0.00406	0.01413	0.03006
## 42	3.6	10	0	1	1.00000	1.00000	0.03384	0.18396	0.63945	1.00000
## 43	4.0	9	1	0	0.88889	0.00711	0.00000	0.00155	0.00406	0.01015
## 44	4.3	8	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 45	4.3	7	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 46	5.0	6	0	1	1.00000	1.00000	0.04773	0.21847	0.57180	1.00000
## 47	6.2	5	1	0	0.80000	0.00003	0.00000	0.00001	0.00001	0.00006
## 48	7.0	4	1	0	0.75000	0.00000	0.00000	0.00000	0.00000	0.00000
## 49	7.5	3	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000
## 50	7.6	2	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000
## 51	9.3	1	0	1	1.00000	1.00000	0.18106	0.42552	0.16601	1.00000

S(t) por grupos de edad

```
aux<- larynx %>%
  mutate("Grupo"=cut(larynx$age,breaks=c(40,55,70,Inf)))
age_fit <- survfit(Surv_general ~ Grupo, type = "kaplan-meier", conf.int = 0.95,
                  conf.type = "plain",data=aux)
ggsurvplot(age_fit,data = aux,palette = pal_lancet()(9),censor.size=7,
            censor.shape= 124,conf.int = TRUE,
            xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(age_fit, data = aux, palette = pal_lancet()(9), censor.size=7,
  censor.shape= 124, conf.int = FALSE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
for(i in levels(aux$Grupo)){
  print(paste("Para el grupo de edad del intervalo:",i))
  delta_aux=(aux$delta)[aux$Grupo==i]
  time_aux=(aux$time)[aux$Grupo==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "Para el grupo de edad del intervalo: (40,55]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 21  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      1.3 21  1  0      0.95238 0.95238 0.00216 0.04647 0.86130 1.00000
## 3      2.4 20  1  0      0.95000 0.90250 0.00408 0.06390 0.77727 1.00000
## 4      3.2 19  0  1      1.00000 1.00000 0.00501 0.07080 0.86124 1.00000
## 5      3.5 18  1  0      0.94444 0.79562 0.00524 0.07240 0.65372 0.93752
## 6      4.0 17  1  0      0.94118 0.73851 0.00652 0.08075 0.58023 0.89678
## 7      4.5 16  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 8      5.9 15  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 9      3.3 14  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 10     4.3 13  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 11     7.5 12  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 12     7.6 11  0  1      1.00000 1.00000 0.01196 0.10935 0.78568 1.00000
## 13     0.3 10  1  0      0.90000 0.28243 0.00184 0.04290 0.19836 0.36650
## 14     1.0  9  1  0      0.88889 0.21628 0.00173 0.04158 0.13479 0.29777
## 15     1.9  8  1  0      0.87500 0.15421 0.00130 0.03610 0.08345 0.22497
## 16     3.2  7  1  0      0.85714 0.09904 0.00077 0.02777 0.04461 0.15347
## 17     3.7  6  0  1      1.00000 1.00000 0.07862 0.28040 0.45043 1.00000
```

```

## 18    4.8  5  0  1    1.00000 1.00000 0.07862 0.28040 0.45043 1.00000
## 19    5.0  4  0  1    1.00000 1.00000 0.07862 0.28040 0.45043 1.00000
## 20   10.1  3  0  1    1.00000 1.00000 0.07862 0.28040 0.45043 1.00000
## 21    1.0  2  1  0    0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 22    4.3  1  0  1    1.00000 1.00000 0.57862 0.76067 0.00000 1.00000
## [1] "Para el grupo de edad del intervalo: (55,70]"
##      Tiempo nj dj cj dif_ponderada    s.t.    Var    Se    L    U
## 1      0.0 41  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      2.5 41  0  1      1.00000 1.00000 0.00000 0.00000 1.00000 1.00000
## 3      3.2 40  1  0      0.97500 0.95062 0.00058 0.02407 0.90345 0.99780
## 4      3.3 39  0  1      1.00000 1.00000 0.00064 0.02532 0.95038 1.00000
## 5      3.5 38  1  0      0.97368 0.89882 0.00109 0.03305 0.83404 0.96360
## 6      4.0 37  1  0      0.97297 0.87197 0.00160 0.03999 0.79360 0.95035
## 7      4.5 36  0  1      1.00000 1.00000 0.00210 0.04586 0.91012 1.00000
## 8      5.5 35  0  1      1.00000 1.00000 0.00210 0.04586 0.91012 1.00000
## 9      5.9 34  0  1      1.00000 1.00000 0.00210 0.04586 0.91012 1.00000
## 10     6.2 33  0  1      1.00000 1.00000 0.00210 0.04586 0.91012 1.00000
## 11     6.5 32  1  0      0.96875 0.72798 0.00165 0.04060 0.64839 0.80756
## 12     6.7 31  0  1      1.00000 1.00000 0.00311 0.05578 0.89068 1.00000
## 13     7.0 30  0  1      1.00000 1.00000 0.00311 0.05578 0.89068 1.00000
## 14     7.4 29  1  0      0.96552 0.63370 0.00174 0.04176 0.55185 0.71554
## 15     8.1 28  0  1      1.00000 1.00000 0.00434 0.06590 0.87084 1.00000
## 16     9.6 27  0  1      1.00000 1.00000 0.00434 0.06590 0.87084 1.00000
## 17    10.7 26  0  1      1.00000 1.00000 0.00434 0.06590 0.87084 1.00000
## 18     1.8 25  1  0      0.96000 0.49959 0.00150 0.03873 0.42368 0.57549
## 19     2.0 24  1  0      0.95833 0.46483 0.00169 0.04111 0.38426 0.54541
## 20     2.6 23  0  1      1.00000 1.00000 0.00782 0.08844 0.82667 1.00000
## 21     3.6 22  1  0      0.95455 0.39440 0.00155 0.03941 0.31715 0.47164
## 22     4.3 21  0  1      1.00000 1.00000 0.00999 0.09993 0.80415 1.00000
## 23     5.0 20  0  1      1.00000 1.00000 0.00999 0.09993 0.80415 1.00000
## 24     7.0 19  1  0      0.94737 0.28836 0.00107 0.03276 0.22415 0.35257
## 25     9.3 18  0  1      1.00000 1.00000 0.01291 0.11362 0.77731 1.00000
## 26     0.5 17  1  0      0.94118 0.21967 0.00080 0.02829 0.16422 0.27512
## 27     1.3 16  1  0      0.93750 0.18675 0.00072 0.02690 0.13402 0.23947
## 28     1.6 15  1  0      0.93333 0.15524 0.00061 0.02480 0.10664 0.20384
## 29     4.5 14  0  1      1.00000 1.00000 0.02551 0.15973 0.68693 1.00000
## 30     4.8 13  0  1      1.00000 1.00000 0.02551 0.15973 0.68693 1.00000
## 31     5.0 12  1  0      0.91667 0.07351 0.00018 0.01337 0.04730 0.09972
## 32     5.1 11  0  1      1.00000 1.00000 0.03309 0.18191 0.64347 1.00000
## 33     6.3 10  1  0      0.90000 0.03434 0.00005 0.00722 0.02019 0.04849
## 34     6.4  9  1  0      0.88889 0.02051 0.00002 0.00494 0.01082 0.03020
## 35     6.5  8  0  1      1.00000 1.00000 0.05809 0.24102 0.52761 1.00000
## 36     7.8  7  1  0      0.85714 0.00454 0.00000 0.00130 0.00199 0.00708
## 37     9.3  6  0  1      1.00000 1.00000 0.08190 0.28618 0.43910 1.00000
## 38     0.1  5  1  0      0.80000 0.00026 0.00000 0.00009 0.00007 0.00044
## 39     0.8  4  1  0      0.75000 0.00002 0.00000 0.00001 0.00000 0.00003
## 40     1.5  3  1  0      0.66667 0.00000 0.00000 0.00000 0.00000 0.00000
## 41     2.0  2  1  0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 42     2.3  1  1  0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
## [1] "Para el grupo de edad del intervalo: (70,Inf]"
##      Tiempo nj dj cj dif_ponderada    s.t.    Var    Se    L    U
## 1      0.0 28  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 28  1  0      0.96429 0.96429 0.00123 0.03507 0.89555 1.00000
## 3      3.3 27  1  0      0.96296 0.92730 0.00236 0.04860 0.83204 1.00000

```

```
## 4      4.3 26 1 0      0.96154 0.88900 0.00339 0.05820 0.77493 1.00000
## 5      5.3 25 1 0      0.96000 0.84935 0.00429 0.06553 0.72091 0.97778
## 6      6.0 24 1 0      0.95833 0.80832 0.00507 0.07122 0.66872 0.94792
## 7      6.1 23 0 1      1.00000 1.00000 0.00776 0.08811 0.82730 1.00000
## 8      6.4 22 1 0      0.95455 0.72207 0.00518 0.07195 0.58105 0.86308
## 9      6.5 21 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 10     7.4 20 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 11     8.1 19 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 12     0.2 18 1 0      0.94444 0.53326 0.00375 0.06126 0.41320 0.65333
## 13     2.2 17 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 14     3.6 16 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 15     4.0 15 1 0      0.93333 0.38064 0.00260 0.05101 0.28066 0.48062
## 16     6.2 14 1 0      0.92857 0.32903 0.00254 0.05039 0.23027 0.42778
## 17     0.3 13 1 0      0.92308 0.27785 0.00231 0.04801 0.18374 0.37195
## 18     0.7 12 1 0      0.91667 0.22782 0.00194 0.04408 0.14142 0.31422
## 19     0.8 11 1 0      0.90909 0.17986 0.00151 0.03880 0.10382 0.25590
## 20     1.8 10 1 0      0.90000 0.13509 0.00105 0.03243 0.07152 0.19865
## 21     1.9 9 1 0      0.88889 0.09483 0.00064 0.02536 0.04512 0.14454
## 22     3.5 8 1 0      0.87500 0.06056 0.00033 0.01811 0.02507 0.09604
## 23     8.0 7 0 1      1.00000 1.00000 0.08939 0.29898 0.41402 1.00000
## 24     0.3 6 1 0      0.83333 0.01509 0.00003 0.00529 0.00473 0.02546
## 25     0.4 5 1 0      0.80000 0.00472 0.00000 0.00196 0.00088 0.00857
## 26     0.8 4 1 0      0.75000 0.00075 0.00000 0.00038 0.00001 0.00150
## 27     2.9 3 0 1      1.00000 1.00000 0.25605 0.50602 0.00822 1.00000
## 28     3.6 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 29     3.8 1 1 0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
```

```
aux%>%
  group_by(Grupo) %>%
  summarise(n=n(), porc=n/10)
```

```
## # A tibble: 3 x 3
##   Grupo      n porc
##   <fct>   <int> <dbl>
## 1 (40,55]    21  23.3
## 2 (55,70]    41  45.6
## 3 (70,Inf]   28  31.1
```

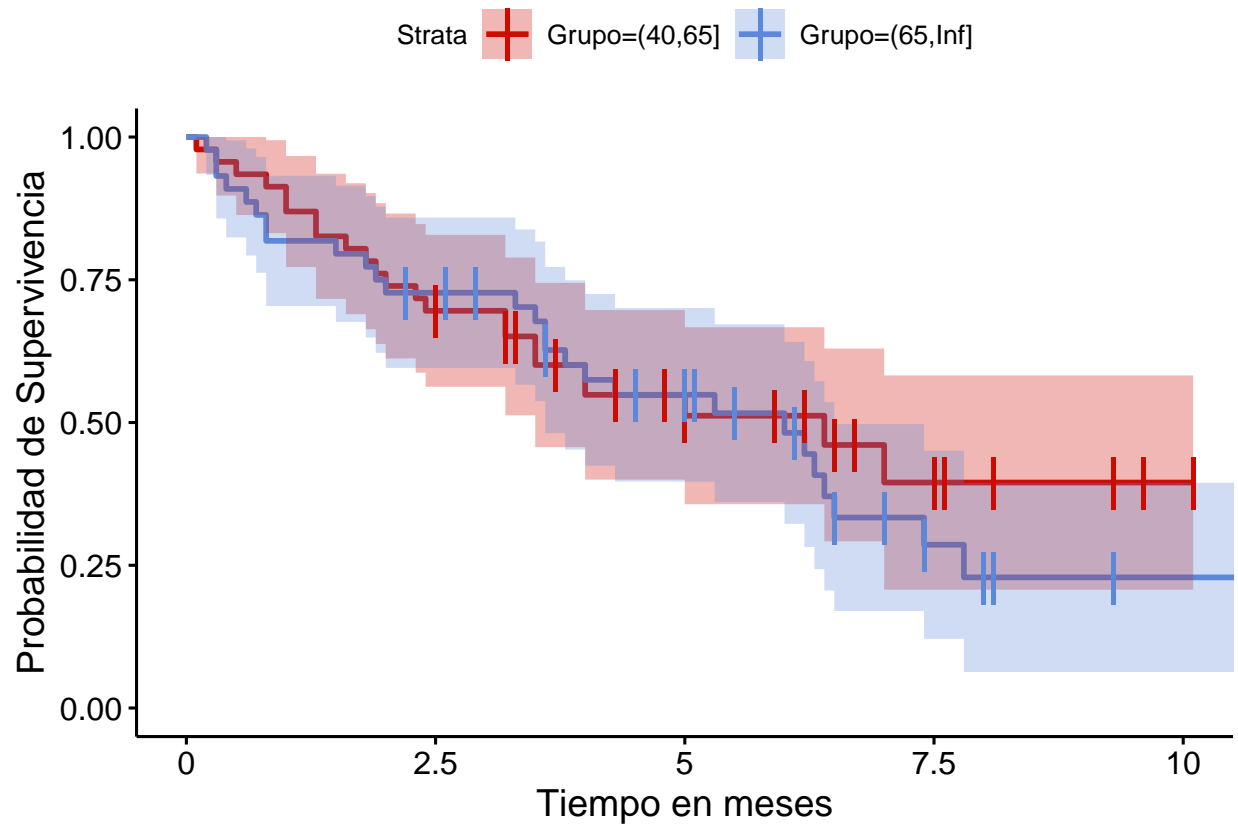
Elegimos esta división por la distribución de los datos, para tratar de compensar los pocos datos que hay en los extremos respecto de la inicial. Además de que graficando en los grupos por edades de 40 a 50, de 51 a 60, de 61 a 70 y de 71 en adelante no obteníamos una gráfica donde se pudieran distinguir muy bien. El intervalo es de 15 años cada grupo, que parece razonable, además de que todos tienen arriba de 20 observaciones.

Bajo esta división, parece que tendríamos que juntar las edades de 40 a 55 y de 56 a 70, ya que parecen tener una función de supervivencia similar, están muy cerca la una de la otra hasta cerca de los 6 meses, que es cuando se da la separación. Por otro lado, la supervivencia del grupo mayor a 70 años parece tener una ruta distinta de las dos anteriores, siendo distinguible.

¿Qué pasará si dividimos entre “Adultos” y “Tercera edad”?

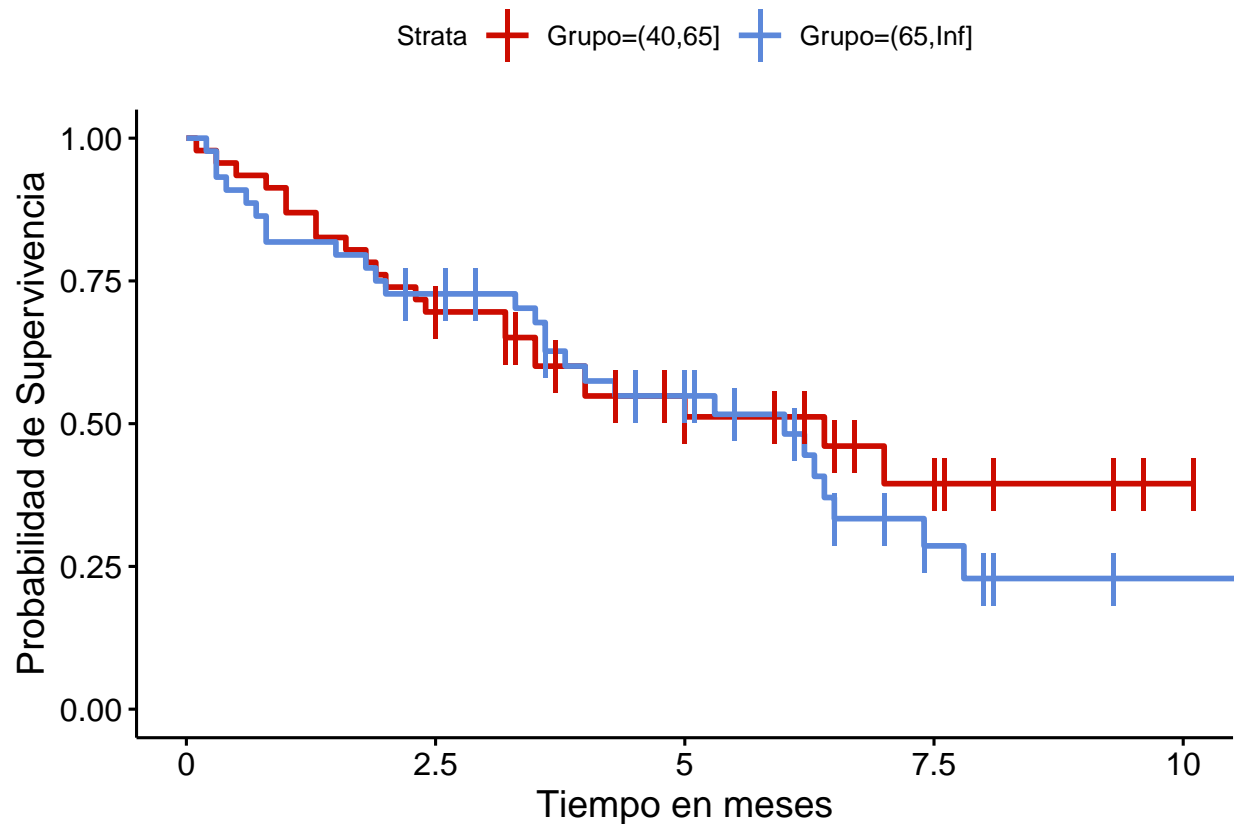
```
aux2<- larynx %>%
  mutate("Grupo"=cut(larynx$age,breaks=c(40,65,Inf)))
age_fit2 <- survfit(Surv_general ~ Grupo, type = "kaplan-meier", conf.int = 0.95,
  conf.type = "plain",data=aux2)
ggsurvplot(age_fit2,data = aux2,palette = pal_startrek()(9),censor.size=7,
  censor.shape= 124,conf.int = TRUE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```

```
## Warning: This manual palette can handle a maximum of 7 values. You have supplied
## 9.
```



```
ggsurvplot(age_fit2,data = aux2,palette = pal_startrek()(9),censor.size=7,
  censor.shape= 124,conf.int = FALSE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```

```
## Warning: This manual palette can handle a maximum of 7 values. You have supplied
## 9.
```



```
for(i in unique(aux2$Grupo)){
  print(i)
  print(paste("Para el grupo de edad del intervalo:",i))
  delta_aux=(aux2$delta)[aux2$Grupo==i]
  time_aux=(aux2$time)[aux2$Grupo==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "(65,Inf]"
## [1] "Para el grupo de edad del intervalo: (65,Inf]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 44  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 44  1  0      0.97727 0.97727 0.00050 0.02247 0.93324 1.00000
## 3      3.3 43  1  0      0.97674 0.95403 0.00099 0.03139 0.89252 1.00000
## 4      4.3 42  1  0      0.97619 0.93026 0.00144 0.03794 0.85591 1.00000
## 5      4.5 41  0  1      1.00000 1.00000 0.00166 0.04078 0.92007 1.00000
## 6      5.3 40  1  0      0.97500 0.88110 0.00179 0.04229 0.79820 0.96399
## 7      5.5 39  0  1      1.00000 1.00000 0.00230 0.04800 0.90592 1.00000
## 8      6.0 38  1  0      0.97368 0.82971 0.00208 0.04556 0.74041 0.91901
## 9      6.1 37  0  1      1.00000 1.00000 0.00302 0.05491 0.89238 1.00000
## 10     6.4 36  1  0      0.97222 0.77605 0.00229 0.04789 0.68218 0.86992
## 11     6.5 35  1  0      0.97143 0.74836 0.00260 0.05103 0.64835 0.84837
## 12     6.5 34  0  1      1.00000 1.00000 0.00465 0.06819 0.86636 1.00000
## 13     7.0 33  0  1      1.00000 1.00000 0.00465 0.06819 0.86636 1.00000
## 14     7.4 32  1  0      0.96875 0.66184 0.00248 0.04978 0.56427 0.75941
## 15     7.4 31  0  1      1.00000 1.00000 0.00566 0.07521 0.85258 1.00000
```



```

## 16      8.1 30 0 1      1.00000 1.00000 0.00566 0.07521 0.85258 1.00000
## 17     10.7 29 0 1      1.00000 1.00000 0.00566 0.07521 0.85258 1.00000
## 18      0.2 28 1 0      0.96429 0.53889 0.00203 0.04502 0.45064 0.62713
## 19      2.2 27 0 1      1.00000 1.00000 0.00698 0.08355 0.83625 1.00000
## 20      2.6 26 0 1      1.00000 1.00000 0.00698 0.08355 0.83625 1.00000
## 21      3.6 25 1 0      0.96000 0.44200 0.00169 0.04110 0.36145 0.52256
## 22      3.6 24 0 1      1.00000 1.00000 0.00865 0.09299 0.81775 1.00000
## 23      4.0 23 1 0      0.95652 0.37609 0.00150 0.03876 0.30011 0.45206
## 24      5.0 22 0 1      1.00000 1.00000 0.01062 0.10307 0.79799 1.00000
## 25      6.2 21 1 0      0.95238 0.31007 0.00125 0.03536 0.24077 0.37937
## 26      0.3 20 1 0      0.95000 0.27739 0.00120 0.03469 0.20941 0.34537
## 27      0.7 19 1 0      0.94737 0.24518 0.00112 0.03340 0.17972 0.31065
## 28      0.8 18 1 0      0.94444 0.21368 0.00100 0.03157 0.15180 0.27555
## 29      1.8 17 1 0      0.94118 0.18314 0.00086 0.02925 0.12582 0.24047
## 30      1.9 16 1 0      0.93750 0.15387 0.00070 0.02651 0.10193 0.20582
## 31      3.5 15 1 0      0.93333 0.12621 0.00055 0.02342 0.08031 0.17212
## 32      4.5 14 0 1      1.00000 1.00000 0.03443 0.18556 0.63631 1.00000
## 33      5.1 13 0 1      1.00000 1.00000 0.03443 0.18556 0.63631 1.00000
## 34      6.3 12 1 0      0.91667 0.05662 0.00013 0.01160 0.03388 0.07937
## 35      7.8 11 1 0      0.90909 0.03914 0.00008 0.00885 0.02180 0.05648
## 36      8.0 10 0 1      1.00000 1.00000 0.05110 0.22605 0.55695 1.00000
## 37      9.3  9 0 1      1.00000 1.00000 0.05110 0.22605 0.55695 1.00000
## 38      0.3  8 1 0      0.87500 0.00715 0.00000 0.00188 0.00347 0.01083
## 39      0.4  7 1 0      0.85714 0.00286 0.00000 0.00087 0.00115 0.00456
## 40      0.8  6 1 0      0.83333 0.00082 0.00000 0.00029 0.00025 0.00138
## 41      1.5  5 1 0      0.80000 0.00013 0.00000 0.00006 0.00002 0.00024
## 42      2.0  4 1 0      0.75000 0.00001 0.00000 0.00000 0.00000 0.00002
## 43      2.9  3 0 1      1.00000 1.00000 0.25943 0.50935 0.00170 1.00000
## 44      3.6  2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 45      3.8  1 1 0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
## [1] "(40,65]"
## [1] "Para el grupo de edad del intervalo: (40,65]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 46 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      1.3 46 1 0      0.97826 0.97826 0.00046 0.02150 0.93612 1.00000
## 3      2.4 45 1 0      0.97778 0.95605 0.00090 0.03005 0.89715 1.00000
## 4      2.5 44 0 1      1.00000 1.00000 0.00099 0.03143 0.93839 1.00000
## 5      3.2 43 1 0      0.97674 0.91017 0.00128 0.03574 0.84012 0.98022
## 6      3.2 42 0 1      1.00000 1.00000 0.00154 0.03927 0.92304 1.00000
## 7      3.3 41 0 1      1.00000 1.00000 0.00154 0.03927 0.92304 1.00000
## 8      3.5 40 1 0      0.97500 0.83759 0.00153 0.03913 0.76089 0.91429
## 9      3.5 39 1 0      0.97436 0.81237 0.00189 0.04343 0.72725 0.89748
## 10     4.0 38 1 0      0.97368 0.78662 0.00221 0.04699 0.69451 0.87872
## 11     4.0 37 1 0      0.97297 0.76034 0.00250 0.04997 0.66240 0.85828
## 12     4.5 36 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 13     5.9 35 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 14     5.9 34 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 15     6.2 33 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 16     6.7 32 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 17     8.1 31 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 18     9.6 30 0 1      1.00000 1.00000 0.00432 0.06572 0.87118 1.00000
## 19     1.8 29 1 0      0.96552 0.53172 0.00157 0.03962 0.45407 0.60936
## 20     2.0 28 1 0      0.96429 0.50108 0.00173 0.04154 0.41966 0.58251
## 21     3.3 27 0 1      1.00000 1.00000 0.00687 0.08291 0.83750 1.00000

```

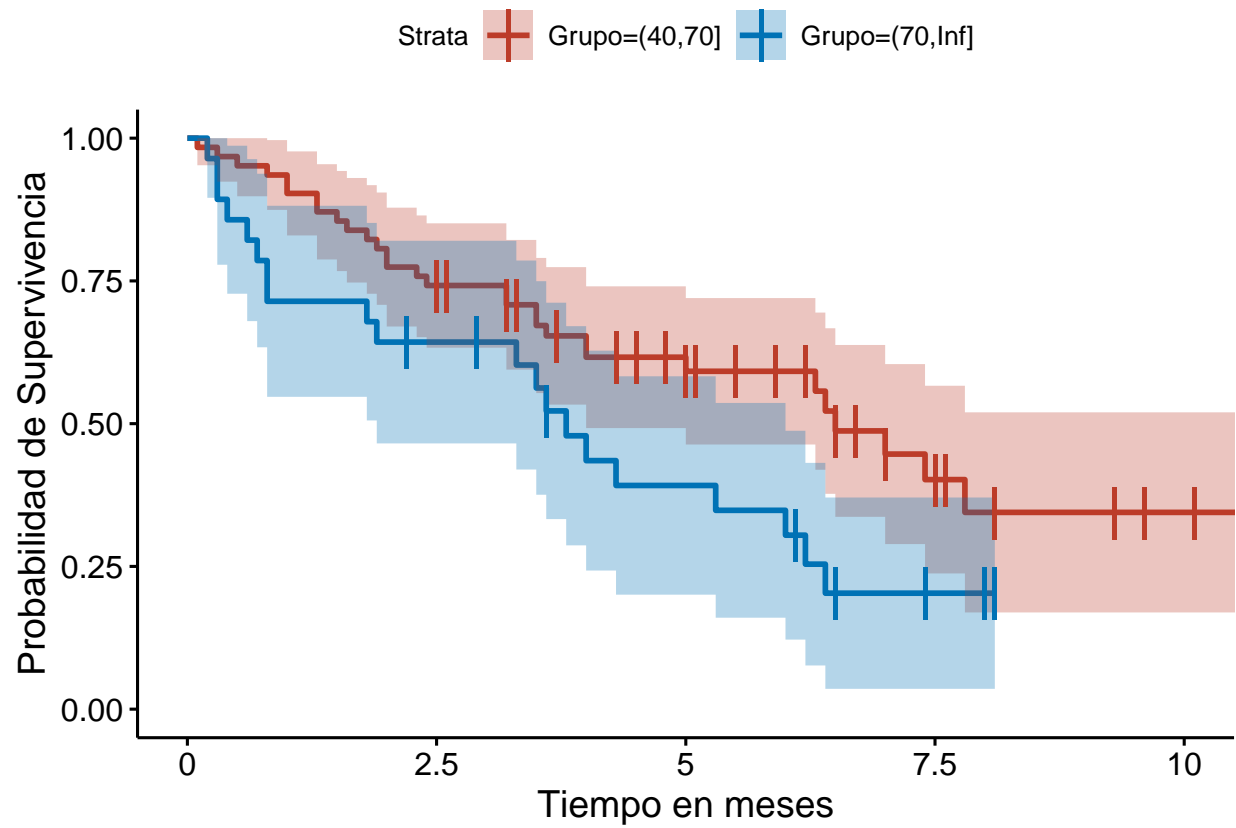
```
## 22 4.3 26 0 1 1.00000 1.00000 0.00687 0.08291 0.83750 1.00000
## 23 4.3 25 0 1 1.00000 1.00000 0.00687 0.08291 0.83750 1.00000
## 24 7.0 24 1 0 0.95833 0.37574 0.00123 0.03502 0.30710 0.44437
## 25 7.5 23 0 1 1.00000 1.00000 0.00869 0.09320 0.81734 1.00000
## 26 7.6 22 0 1 1.00000 1.00000 0.00869 0.09320 0.81734 1.00000
## 27 9.3 21 0 1 1.00000 1.00000 0.00869 0.09320 0.81734 1.00000
## 28 0.3 20 1 0 0.95000 0.25034 0.00071 0.02663 0.19815 0.30254
## 29 0.5 19 1 0 0.94737 0.22005 0.00069 0.02626 0.16858 0.27152
## 30 1.0 18 1 0 0.94444 0.19060 0.00064 0.02522 0.14117 0.24003
## 31 1.3 17 1 0 0.94118 0.16223 0.00056 0.02361 0.11595 0.20851
## 32 1.6 16 1 0 0.93750 0.13524 0.00046 0.02153 0.09304 0.17745
## 33 1.9 15 1 0 0.93333 0.10995 0.00036 0.01908 0.07255 0.14734
## 34 3.2 14 1 0 0.92857 0.08668 0.00027 0.01636 0.05462 0.11873
## 35 3.7 13 0 1 1.00000 1.00000 0.03561 0.18870 0.63015 1.00000
## 36 4.8 12 0 1 1.00000 1.00000 0.03561 0.18870 0.63015 1.00000
## 37 4.8 11 0 1 1.00000 1.00000 0.03561 0.18870 0.63015 1.00000
## 38 5.0 10 1 0 0.90000 0.02028 0.00002 0.00438 0.01169 0.02887
## 39 5.0 9 0 1 1.00000 1.00000 0.04672 0.21615 0.57636 1.00000
## 40 6.4 8 1 0 0.87500 0.00547 0.00000 0.00139 0.00275 0.00820
## 41 6.5 7 0 1 1.00000 1.00000 0.06458 0.25412 0.50193 1.00000
## 42 10.1 6 0 1 1.00000 1.00000 0.06458 0.25412 0.50193 1.00000
## 43 0.1 5 1 0 0.80000 0.00009 0.00000 0.00003 0.00003 0.00014
## 44 0.8 4 1 0 0.75000 0.00000 0.00000 0.00000 0.00000 0.00001
## 45 1.0 3 1 0 0.66667 0.00000 0.00000 0.00000 0.00000 0.00000
## 46 2.3 2 1 0 0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 47 4.3 1 0 1 1.00000 1.00000 0.86458 0.92983 0.00000 1.00000
```

```
aux2%>%
  group_by(Grupo) %>%
  summarise(n=n(), porc=n/.9)
```

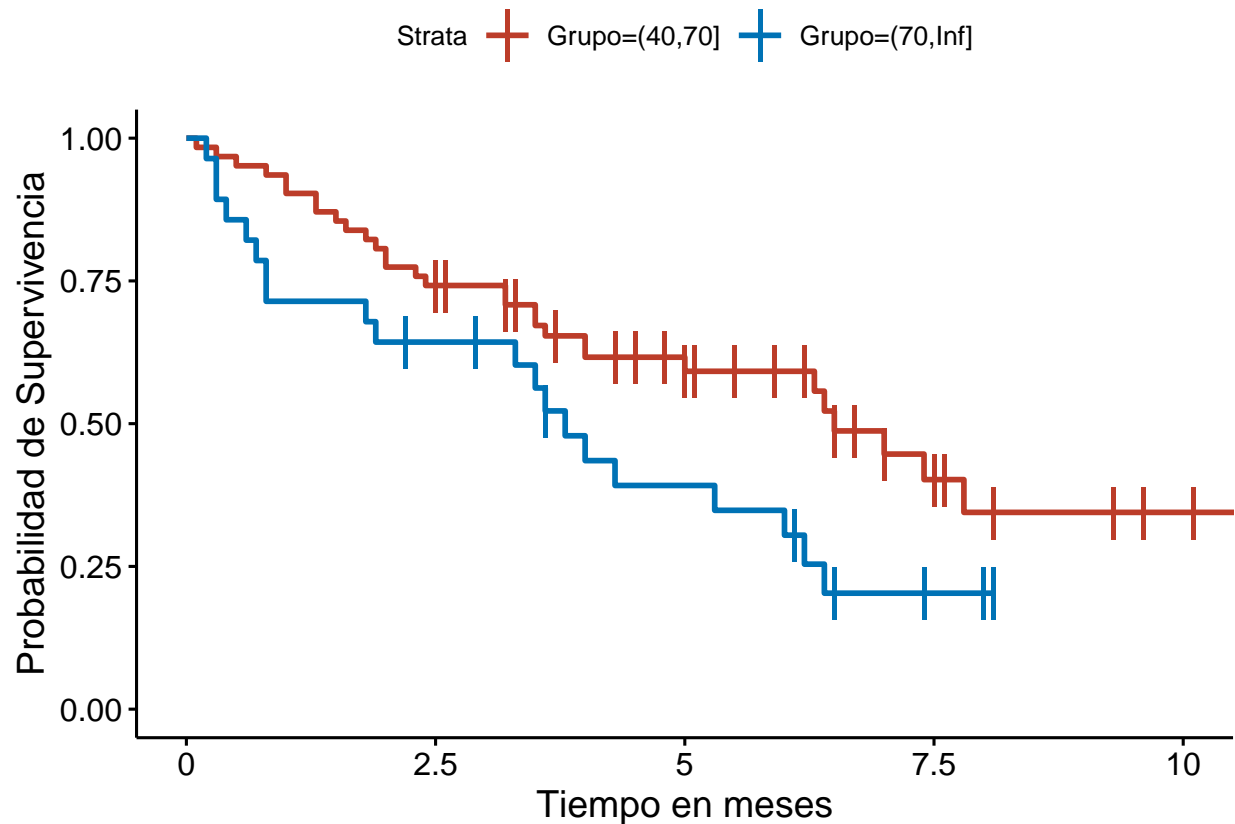
```
## # A tibble: 2 x 3
##   Grupo      n porc
##   <fct>   <int> <dbl>
## 1 (40,65]    46  51.1
## 2 (65,Inf]   44  48.9
```

¡Son indistinguibles! Prácticamente son la misma función de supervivencia, a simple vista. Aquí hay casi la misma cantidad de datos en cada categoría y la división suena la más natural, puesto que es la división que haría INEGI para adultos y adultos mayores.

```
aux3<- larynx %>%
  mutate("Grupo"=cut(larynx$age,breaks=c(40,70,Inf)))
age_fit3 <- survfit(Surv_general ~ Grupo, type = "kaplan-meier", conf.int = 0.95,
  conf.type = "plain",data=aux3)
ggsurvplot(age_fit3,data = aux3,palette = pal_nejm()(8),censor.size=7,
  censor.shape= 124,conf.int = TRUE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(age_fit3, data = aux3, palette = pal_nejm()(8), censor.size=7,
  censor.shape= 124, conf.int = FALSE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
for(i in levels(aux3$Grupo)){
  print(paste("Para el grupo de edad del intervalo:",i))
  delta_aux=(aux3$delta)[aux3$Grupo==i]
  time_aux=(aux3$time)[aux3$Grupo==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "Para el grupo de edad del intervalo: (40,70]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 62  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      1.3 62  1  0      0.98387 0.98387 0.00026 0.01600 0.95251 1.00000
## 3      2.4 61  1  0      0.98361 0.96748 0.00050 0.02243 0.92351 1.00000
## 4      2.5 60  0  1      1.00000 1.00000 0.00054 0.02319 0.95455 1.00000
## 5      3.2 59  1  0      0.98305 0.93391 0.00072 0.02690 0.88118 0.98664
## 6      3.2 58  0  1      1.00000 1.00000 0.00083 0.02881 0.94354 1.00000
## 7      3.3 57  0  1      1.00000 1.00000 0.00083 0.02881 0.94354 1.00000
## 8      3.5 56  1  0      0.98214 0.88150 0.00090 0.02995 0.82280 0.94021
## 9      3.5 55  1  0      0.98182 0.86347 0.00111 0.03334 0.79812 0.92883
## 10     4.0 54  1  0      0.98148 0.84516 0.00131 0.03626 0.77409 0.91623
## 11     4.0 53  1  0      0.98113 0.82656 0.00151 0.03880 0.75051 0.90261
## 12     4.5 52  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
## 13     4.5 51  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
## 14     5.5 50  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
## 15     5.9 49  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
## 16     5.9 48  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
## 17     6.2 47  0  1      1.00000 1.00000 0.00220 0.04694 0.90800 1.00000
```

```

## 18    6.5 46 1 0    0.97826 0.68822 0.00127 0.03567 0.61831 0.75814
## 19    6.7 45 0 1    1.00000 1.00000 0.00269 0.05183 0.89841 1.00000
## 20    7.0 44 0 1    1.00000 1.00000 0.00269 0.05183 0.89841 1.00000
## 21    7.4 43 1 0    0.97674 0.62462 0.00126 0.03556 0.55493 0.69431
## 22    8.1 42 0 1    1.00000 1.00000 0.00324 0.05692 0.88843 1.00000
## 23    9.6 41 0 1    1.00000 1.00000 0.00324 0.05692 0.88843 1.00000
## 24   10.7 40 0 1    1.00000 1.00000 0.00324 0.05692 0.88843 1.00000
## 25    1.8 39 1 0    0.97436 0.53611 0.00113 0.03354 0.47037 0.60186
## 26    2.0 38 1 0    0.97368 0.51340 0.00122 0.03492 0.44496 0.58184
## 27    2.6 37 0 1    1.00000 1.00000 0.00463 0.06802 0.86669 1.00000
## 28    3.3 36 0 1    1.00000 1.00000 0.00463 0.06802 0.86669 1.00000
## 29    3.6 35 1 0    0.97143 0.44412 0.00108 0.03284 0.37976 0.50848
## 30    4.3 34 0 1    1.00000 1.00000 0.00547 0.07394 0.85509 1.00000
## 31    4.3 33 0 1    1.00000 1.00000 0.00547 0.07394 0.85509 1.00000
## 32    5.0 32 0 1    1.00000 1.00000 0.00547 0.07394 0.85509 1.00000
## 33    7.0 31 1 0    0.96774 0.35019 0.00080 0.02832 0.29468 0.40571
## 34    7.5 30 0 1    1.00000 1.00000 0.00654 0.08088 0.84147 1.00000
## 35    7.6 29 0 1    1.00000 1.00000 0.00654 0.08088 0.84147 1.00000
## 36    9.3 28 0 1    1.00000 1.00000 0.00654 0.08088 0.84147 1.00000
## 37    0.3 27 1 0    0.96296 0.25701 0.00053 0.02294 0.21205 0.30197
## 38    0.5 26 1 0    0.96154 0.23430 0.00052 0.02284 0.18953 0.27907
## 39    1.0 25 1 0    0.96000 0.21199 0.00050 0.02241 0.16807 0.25590
## 40    1.3 24 1 0    0.95833 0.19017 0.00047 0.02167 0.14770 0.23264
## 41    1.6 23 1 0    0.95652 0.16896 0.00043 0.02067 0.12846 0.20947
## 42    1.9 22 1 0    0.95455 0.14848 0.00038 0.01943 0.11040 0.18656
## 43    3.2 21 1 0    0.95238 0.12884 0.00032 0.01799 0.09357 0.16411
## 44    3.7 20 0 1    1.00000 1.00000 0.01950 0.13966 0.72627 1.00000
## 45    4.5 19 0 1    1.00000 1.00000 0.01950 0.13966 0.72627 1.00000
## 46    4.8 18 0 1    1.00000 1.00000 0.01950 0.13966 0.72627 1.00000
## 47    4.8 17 0 1    1.00000 1.00000 0.01950 0.13966 0.72627 1.00000
## 48    5.0 16 1 0    0.93750 0.04816 0.00005 0.00741 0.03363 0.06268
## 49    5.0 15 0 1    1.00000 1.00000 0.02367 0.15386 0.69845 1.00000
## 50    5.1 14 0 1    1.00000 1.00000 0.02367 0.15386 0.69845 1.00000
## 51    6.3 13 1 0    0.92308 0.01828 0.00001 0.00317 0.01206 0.02449
## 52    6.4 12 1 0    0.91667 0.01182 0.00001 0.00229 0.00733 0.01632
## 53    6.5 11 0 1    1.00000 1.00000 0.03766 0.19406 0.61966 1.00000
## 54    7.8 10 1 0    0.90000 0.00376 0.00000 0.00083 0.00213 0.00538
## 55    9.3  9 0 1    1.00000 1.00000 0.04877 0.22084 0.56717 1.00000
## 56   10.1  8 0 1    1.00000 1.00000 0.04877 0.22084 0.56717 1.00000
## 57    0.1  7 1 0    0.85714 0.00018 0.00000 0.00005 0.00008 0.00027
## 58    0.8  6 1 0    0.83333 0.00003 0.00000 0.00001 0.00001 0.00005
## 59    1.0  5 1 0    0.80000 0.00000 0.00000 0.00000 0.00000 0.00000
## 60    1.5  4 1 0    0.75000 0.00000 0.00000 0.00000 0.00000 0.00000
## 61    2.0  3 1 0    0.66667 0.00000 0.00000 0.00000 0.00000 0.00000
## 62    2.3  2 1 0    0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 63    4.3  1 0 1    1.00000 1.00000 0.90591 0.95179 0.00000 1.00000
## [1] "Para el grupo de edad del intervalo: (70,Inf]"
##      Tiempo nj dj cj dif_ponderada    s.t.      Var      Se      L      U
## 1      0.0 28 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 28 1 0      0.96429 0.96429 0.00123 0.03507 0.89555 1.00000
## 3      3.3 27 1 0      0.96296 0.92730 0.00236 0.04860 0.83204 1.00000
## 4      4.3 26 1 0      0.96154 0.88900 0.00339 0.05820 0.77493 1.00000
## 5      5.3 25 1 0      0.96000 0.84935 0.00429 0.06553 0.72091 0.97778
## 6      6.0 24 1 0      0.95833 0.80832 0.00507 0.07122 0.66872 0.94792

```

```
## 7      6.1 23 0 1      1.00000 1.00000 0.00776 0.08811 0.82730 1.00000
## 8      6.4 22 1 0      0.95455 0.72207 0.00518 0.07195 0.58105 0.86308
## 9      6.5 21 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 10     7.4 20 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 11     8.1 19 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 12     0.2 18 1 0      0.94444 0.53326 0.00375 0.06126 0.41320 0.65333
## 13     2.2 17 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 14     3.6 16 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 15     4.0 15 1 0      0.93333 0.38064 0.00260 0.05101 0.28066 0.48062
## 16     6.2 14 1 0      0.92857 0.32903 0.00254 0.05039 0.23027 0.42778
## 17     0.3 13 1 0      0.92308 0.27785 0.00231 0.04801 0.18374 0.37195
## 18     0.7 12 1 0      0.91667 0.22782 0.00194 0.04408 0.14142 0.31422
## 19     0.8 11 1 0      0.90909 0.17986 0.00151 0.03880 0.10382 0.25590
## 20     1.8 10 1 0      0.90000 0.13509 0.00105 0.03243 0.07152 0.19865
## 21     1.9 9 1 0      0.88889 0.09483 0.00064 0.02536 0.04512 0.14454
## 22     3.5 8 1 0      0.87500 0.06056 0.00033 0.01811 0.02507 0.09604
## 23     8.0 7 0 1      1.00000 1.00000 0.08939 0.29898 0.41402 1.00000
## 24     0.3 6 1 0      0.83333 0.01509 0.00003 0.00529 0.00473 0.02546
## 25     0.4 5 1 0      0.80000 0.00472 0.00000 0.00196 0.00088 0.00857
## 26     0.8 4 1 0      0.75000 0.00075 0.00000 0.00038 0.00001 0.00150
## 27     2.9 3 0 1      1.00000 1.00000 0.25605 0.50602 0.00822 1.00000
## 28     3.6 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 29     3.8 1 1 0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
```

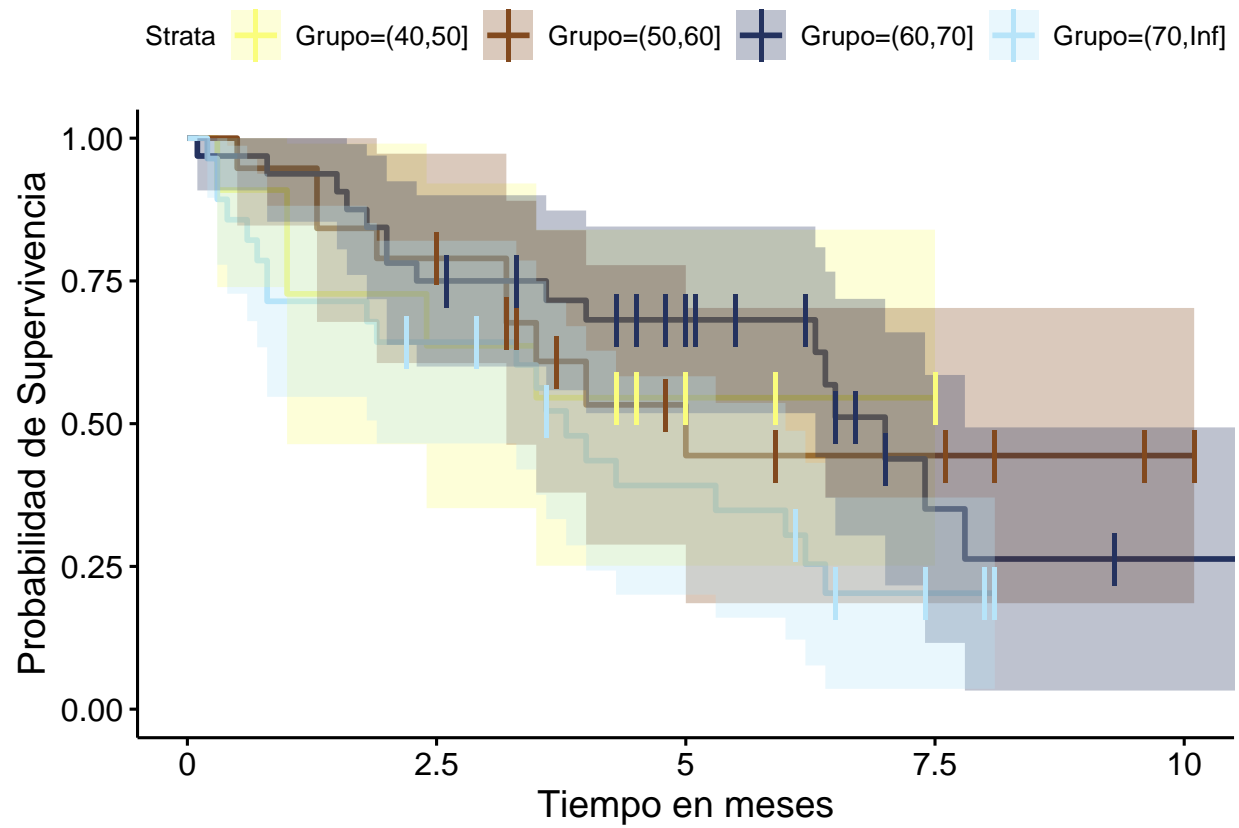
Bajo esta división parecen ser diferentes funciones de supervivencia. Lo confirmaremos en las pruebas Log-Rank. Puede usarse como una nueva clasificación entre “Adultos” y personas en “Vejez”.

```
aux3%>%
  group_by(Grupo) %>%
  summarise(n=n(), porc=n/.9)
```

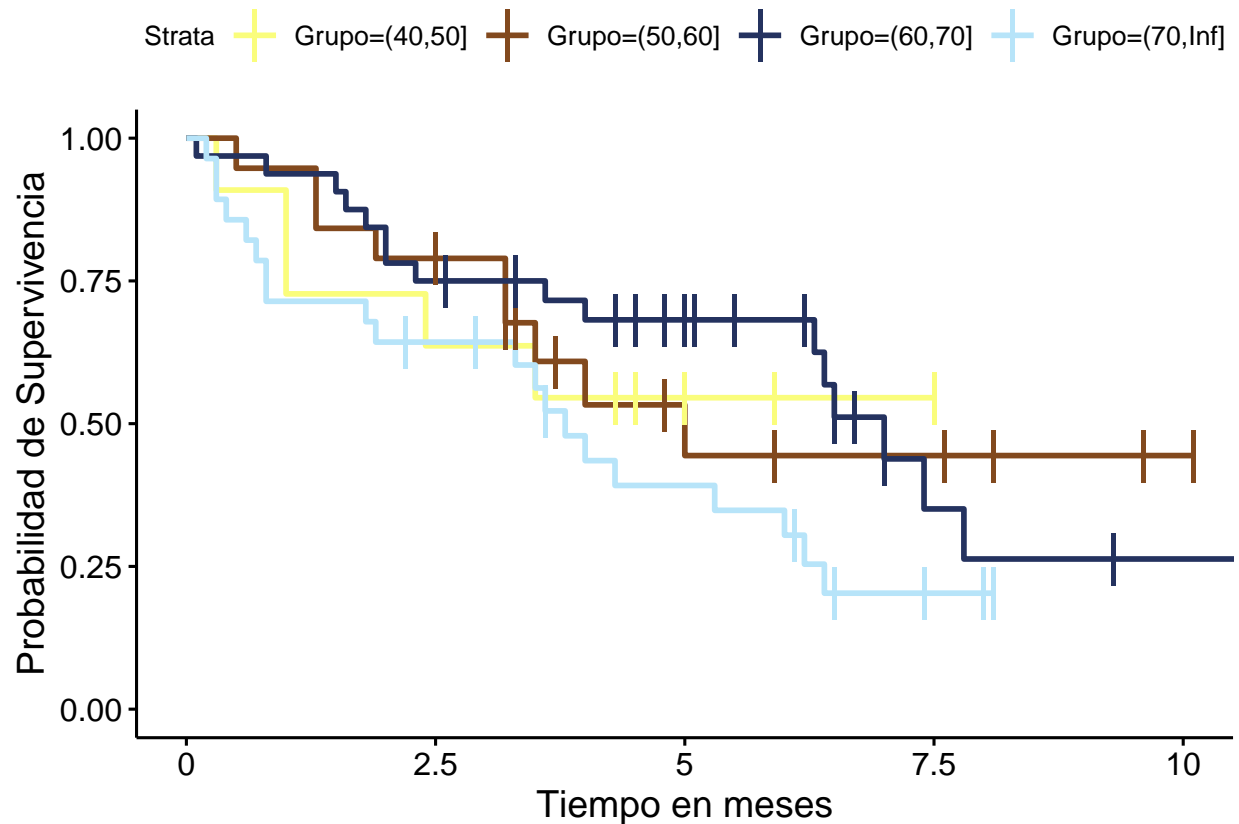
```
## # A tibble: 2 x 3
##   Grupo      n porc
##   <fct>   <int> <dbl>
## 1 (40,70]     62  68.9
## 2 (70,Inf]    28  31.1
```

Hay una diferencia de 34 observaciones de un grupo respecto al otro, pero el segundo grupo todavía tiene una cantidad considerable de obsevaciones, que es el 31% de las observaciones, siendo 28.

```
aux4<- larynx %>%
  mutate("Grupo"=cut(larynx$age,breaks=c(40,50,60,70,Inf)))
age_fit4 <- survfit(Surv_general ~ Grupo, type = "kaplan-meier", conf.int = 0.95,
  conf.type = "plain",data=aux4)
ggsurvplot(age_fit4,data = aux4,palette = pal_rickandmorty()(9),censor.size=7,
  censor.shape= 124,conf.int = TRUE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
ggsurvplot(age_fit4, data = aux4, palette = pal_rickandmarty()(9), censor.size=7,
  censor.shape= 124, conf.int = FALSE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



```
for(i in levels(aux4$Grupo)){
  print(paste("Para el grupo de edad del intervalo:",i))
  delta_aux=(aux4$delta)[aux4$Grupo==i]
  time_aux=(aux4$time)[aux4$Grupo==i]
  print(KM(delta_aux,time_aux))
}
```

```
## [1] "Para el grupo de edad del intervalo: (40,50]"
##      Tiempo nj dj cj dif_ponderada   s.t.      Var      Se      L      U
## 1      0.0 11  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      2.4 11  1  0      0.90909 0.90909 0.00751 0.08668 0.73920 1.00000
## 3      3.5 10  1  0      0.90000 0.81000 0.01325 0.11513 0.58435 1.00000
## 4      4.5  9  0  1      1.00000 1.00000 0.02020 0.14213 0.72142 1.00000
## 5      5.9  8  0  1      1.00000 1.00000 0.02020 0.14213 0.72142 1.00000
## 6      4.3  7  0  1      1.00000 1.00000 0.02020 0.14213 0.72142 1.00000
## 7      7.5  6  0  1      1.00000 1.00000 0.02020 0.14213 0.72142 1.00000
## 8      0.3  5  1  0      0.80000 0.20972 0.00309 0.05557 0.10081 0.31862
## 9      1.0  4  1  0      0.75000 0.10011 0.00154 0.03923 0.02323 0.17700
## 10     5.0  3  0  1      1.00000 1.00000 0.15354 0.39184 0.23202 1.00000
## 11     1.0  2  1  0      0.50000 0.00098 0.00000 0.00079 0.00000 0.00252
## 12     4.3  1  0  1      1.00000 1.00000 0.65354 0.80842 0.00000 1.00000
## [1] "Para el grupo de edad del intervalo: (50,60]"
##      Tiempo nj dj cj dif_ponderada   s.t.      Var      Se      L      U
## 1      0.0 19  0  0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      1.3 19  1  0      0.94737 0.94737 0.00262 0.05123 0.84696 1.00000
## 3      2.5 18  0  1      1.00000 1.00000 0.00292 0.05407 0.89402 1.00000
```



```

## 4      3.2 17 1 0      0.94118 0.83371 0.00459 0.06773 0.70095 0.96646
## 5      3.2 16 0 1      1.00000 1.00000 0.00660 0.08124 0.84077 1.00000
## 6      3.5 15 1 0      0.93333 0.70825 0.00570 0.07549 0.56028 0.85621
## 7      4.0 14 1 0      0.92857 0.64105 0.00693 0.08323 0.47792 0.80418
## 8      5.9 13 0 1      1.00000 1.00000 0.01686 0.12983 0.74553 1.00000
## 9      8.1 12 0 1      1.00000 1.00000 0.01686 0.12983 0.74553 1.00000
## 10     9.6 11 0 1      1.00000 1.00000 0.01686 0.12983 0.74553 1.00000
## 11     3.3 10 0 1      1.00000 1.00000 0.01686 0.12983 0.74553 1.00000
## 12     7.6 9 0 1      1.00000 1.00000 0.01686 0.12983 0.74553 1.00000
## 13     0.5 8 1 0      0.87500 0.20142 0.00141 0.03753 0.12786 0.27497
## 14     1.3 7 1 0      0.85714 0.13480 0.00106 0.03261 0.07089 0.19872
## 15     1.9 6 1 0      0.83333 0.07789 0.00056 0.02361 0.03162 0.12415
## 16     3.2 5 1 0      0.80000 0.03518 0.00018 0.01325 0.00921 0.06116
## 17     3.7 4 0 1      1.00000 1.00000 0.14186 0.37664 0.26180 1.00000
## 18     4.8 3 0 1      1.00000 1.00000 0.14186 0.37664 0.26180 1.00000
## 19     5.0 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00001
## 20    10.1 1 0 1      1.00000 1.00000 0.64186 0.80116 0.00000 1.00000
## [1] "Para el grupo de edad del intervalo: (60,70]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U
## 1      0.0 32 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      3.3 32 0 1      1.00000 1.00000 0.00000 0.00000 1.00000 1.00000
## 3      4.0 31 1 0      0.96774 0.93652 0.00094 0.03071 0.87633 0.99671
## 4      4.5 30 0 1      1.00000 1.00000 0.00108 0.03279 0.93573 1.00000
## 5      5.5 29 0 1      1.00000 1.00000 0.00108 0.03279 0.93573 1.00000
## 6      6.2 28 0 1      1.00000 1.00000 0.00108 0.03279 0.93573 1.00000
## 7      6.5 27 1 0      0.96296 0.79737 0.00159 0.03987 0.71923 0.87550
## 8      6.7 26 0 1      1.00000 1.00000 0.00250 0.05000 0.90201 1.00000
## 9      7.0 25 0 1      1.00000 1.00000 0.00250 0.05000 0.90201 1.00000
## 10     7.4 24 1 0      0.95833 0.68179 0.00200 0.04477 0.59405 0.76953
## 11    10.7 23 0 1      1.00000 1.00000 0.00431 0.06566 0.87131 1.00000
## 12     1.8 22 1 0      0.95455 0.59946 0.00233 0.04824 0.50491 0.69401
## 13     2.0 21 1 0      0.95238 0.55684 0.00275 0.05240 0.45413 0.65955
## 14     2.6 20 0 1      1.00000 1.00000 0.00886 0.09411 0.81555 1.00000
## 15     3.6 19 1 0      0.94737 0.46910 0.00259 0.05092 0.36931 0.56889
## 16     4.3 18 0 1      1.00000 1.00000 0.01178 0.10854 0.78727 1.00000
## 17     5.0 17 0 1      1.00000 1.00000 0.01178 0.10854 0.78727 1.00000
## 18     7.0 16 1 0      0.93750 0.33382 0.00178 0.04216 0.25120 0.41644
## 19     9.3 15 0 1      1.00000 1.00000 0.01595 0.12628 0.75249 1.00000
## 20     1.6 14 1 0      0.92857 0.24462 0.00128 0.03582 0.17441 0.31483
## 21     4.5 13 0 1      1.00000 1.00000 0.02144 0.14643 0.71300 1.00000
## 22     4.8 12 0 1      1.00000 1.00000 0.02144 0.14643 0.71300 1.00000
## 23     5.1 11 0 1      1.00000 1.00000 0.02144 0.14643 0.71300 1.00000
## 24     6.3 10 1 0      0.90000 0.08863 0.00026 0.01599 0.05729 0.11997
## 25     6.4 9 1 0      0.88889 0.05920 0.00016 0.01276 0.03420 0.08421
## 26     6.5 8 0 1      1.00000 1.00000 0.04644 0.21550 0.57762 1.00000
## 27     7.8 7 1 0      0.85714 0.01817 0.00002 0.00482 0.00873 0.02761
## 28     9.3 6 0 1      1.00000 1.00000 0.07025 0.26505 0.48051 1.00000
## 29     0.1 5 1 0      0.80000 0.00193 0.00000 0.00067 0.00062 0.00325
## 30     0.8 4 1 0      0.75000 0.00024 0.00000 0.00011 0.00003 0.00045
## 31     1.5 3 1 0      0.66667 0.00001 0.00000 0.00000 0.00000 0.00001
## 32     2.0 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 33     2.3 1 1 0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
## [1] "Para el grupo de edad del intervalo: (70,Inf]"
##      Tiempo nj dj cj dif_ponderada      s.t.      Var      Se      L      U

```

```
## 1      0.0 28 0 0      1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## 2      0.6 28 1 0      0.96429 0.96429 0.00123 0.03507 0.89555 1.00000
## 3      3.3 27 1 0      0.96296 0.92730 0.00236 0.04860 0.83204 1.00000
## 4      4.3 26 1 0      0.96154 0.88900 0.00339 0.05820 0.77493 1.00000
## 5      5.3 25 1 0      0.96000 0.84935 0.00429 0.06553 0.72091 0.97778
## 6      6.0 24 1 0      0.95833 0.80832 0.00507 0.07122 0.66872 0.94792
## 7      6.1 23 0 1      1.00000 1.00000 0.00776 0.08811 0.82730 1.00000
## 8      6.4 22 1 0      0.95455 0.72207 0.00518 0.07195 0.58105 0.86308
## 9      6.5 21 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 10     7.4 20 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 11     8.1 19 0 1      1.00000 1.00000 0.00993 0.09964 0.80471 1.00000
## 12     0.2 18 1 0      0.94444 0.53326 0.00375 0.06126 0.41320 0.65333
## 13     2.2 17 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 14     3.6 16 0 1      1.00000 1.00000 0.01320 0.11488 0.77485 1.00000
## 15     4.0 15 1 0      0.93333 0.38064 0.00260 0.05101 0.28066 0.48062
## 16     6.2 14 1 0      0.92857 0.32903 0.00254 0.05039 0.23027 0.42778
## 17     0.3 13 1 0      0.92308 0.27785 0.00231 0.04801 0.18374 0.37195
## 18     0.7 12 1 0      0.91667 0.22782 0.00194 0.04408 0.14142 0.31422
## 19     0.8 11 1 0      0.90909 0.17986 0.00151 0.03880 0.10382 0.25590
## 20     1.8 10 1 0      0.90000 0.13509 0.00105 0.03243 0.07152 0.19865
## 21     1.9 9 1 0      0.88889 0.09483 0.00064 0.02536 0.04512 0.14454
## 22     3.5 8 1 0      0.87500 0.06056 0.00033 0.01811 0.02507 0.09604
## 23     8.0 7 0 1      1.00000 1.00000 0.08939 0.29898 0.41402 1.00000
## 24     0.3 6 1 0      0.83333 0.01509 0.00003 0.00529 0.00473 0.02546
## 25     0.4 5 1 0      0.80000 0.00472 0.00000 0.00196 0.00088 0.00857
## 26     0.8 4 1 0      0.75000 0.00075 0.00000 0.00038 0.00001 0.00150
## 27     2.9 3 0 1      1.00000 1.00000 0.25605 0.50602 0.00822 1.00000
## 28     3.6 2 1 0      0.50000 0.00000 0.00000 0.00000 0.00000 0.00000
## 29     3.8 1 1 0      0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
```

```
aux4%>%
  group_by(Grupo) %>%
  summarise(n=n(), porc=n/.9)
```

```
## # A tibble: 4 x 3
##   Grupo      n porc
##   <fct>   <int> <dbl>
## 1 (40,50]    11 12.2
## 2 (50,60]    19 21.1
## 3 (60,70]    32 35.6
## 4 (70,Inf]   28 31.1
```

Esta es la división que comentamos el inicio, además notemos que en el primer grupo tenemos muy pocos datos. Respecto a la gráfica, notemos lo siguiente: Cruzan demasiado entre sí durante los primeros meses, y entre el mes 4 y 5 la supervivencia del grupo de 40 a 50 y de 51 a 60 parece ser la misma, y en los primeros 2 meses y medio, la supervivencia del grupo de 51 a 60 parece la misma que la del grupo de 61 a 70. En el mismo periodo, la supervivencia del grupo de 70 en adelante parece estar muy cerca de la supervivencia del grupo de 40 a 50, lo cual es sorprendente puesto que son los grupos contrarios: El más joven y el más viejo. Ello parece contraintuitivo

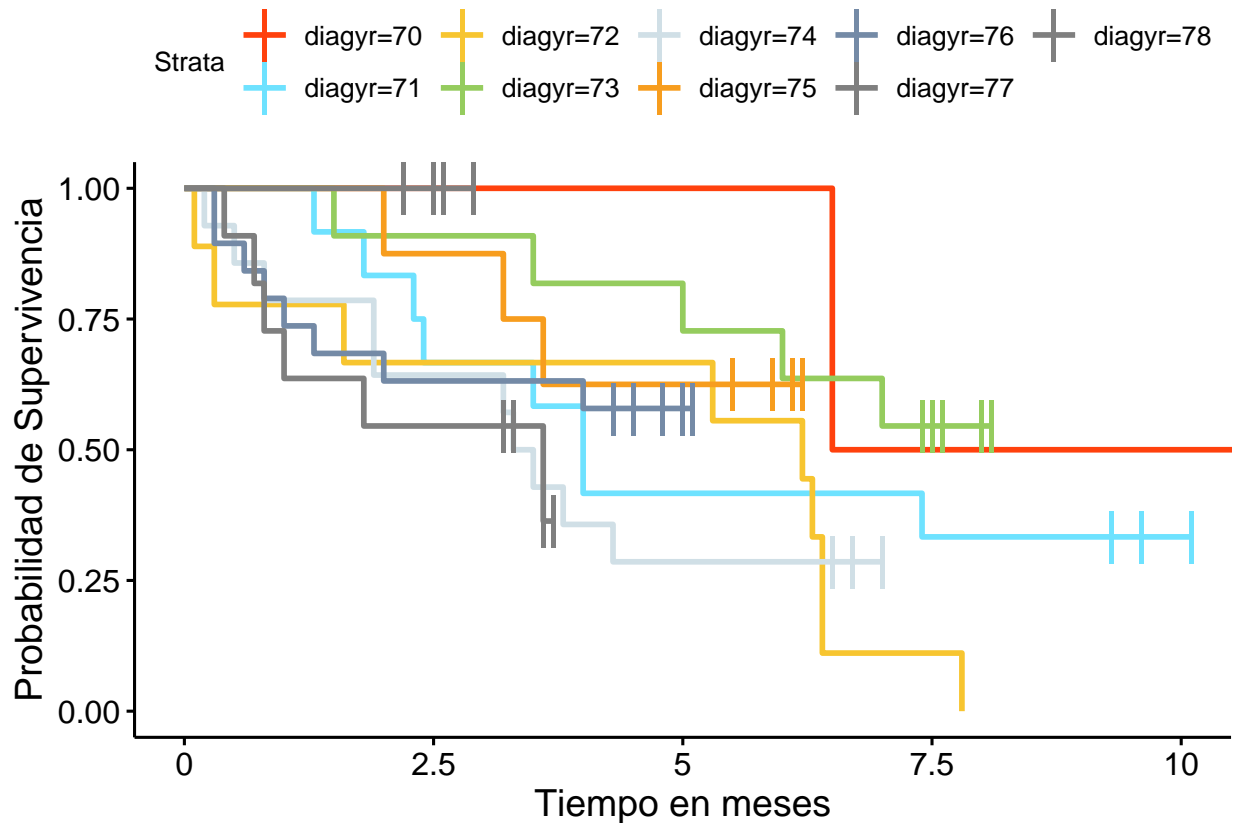
Variables más importantes

Podemos observar de manera gráfica que, definitivamente la edad es un factor que afecta siempre y cuando partamos los grupos de edad de 40 a 70 y de 70 en adelante, así mismo se puede observar que la etapa juega un rol fundamental para el tiempo de supervivencia; esto se comprueba en la última parte de esta tarea.

Veamos qué pasa, de manera gráfica, con las demás variables:

```
year_fit <- survfit(Surv_general ~ diagyr, type = "kaplan-meier", conf.int = 0.95,
                    conf.type = "plain", data=larynx)
ggsurvplot(year_fit, data = larynx, palette = pal_tron()(9), censor.size=7,
            censor.shape= 124, conf.int = FALSE,
            xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia" )
```

```
## Warning: This manual palette can handle a maximum of 7 values. You have supplied
## 9.
```



A simple vista, vemos que hay muchos empates hasta el mes 5, hagamos el test:

```
survdif(Surv(time,delta)~diagyr,rho=0,data=larynx) # log-rank
```

```
## Call:
```

```
## survdiff(formula = Surv(time, delta) ~ diagyr, data = larynx,
##          rho = 0)
```

```
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## diagyr=70 2      1    2.11  0.58591  0.62967
## diagyr=71 12     8    8.49  0.02833  0.03560
## diagyr=72 9      9    5.40  2.40634  2.76979
## diagyr=73 11     5    9.73  2.29988  3.03967
## diagyr=74 14    10    6.88  1.41531  1.68382
## diagyr=75 8      3    4.56  0.53136  0.60491
## diagyr=76 19     8    8.19  0.00444  0.00571
## diagyr=77 11     6    3.33  2.15139  2.42802
## diagyr=78 4      0    1.32  1.31830  1.40479
```

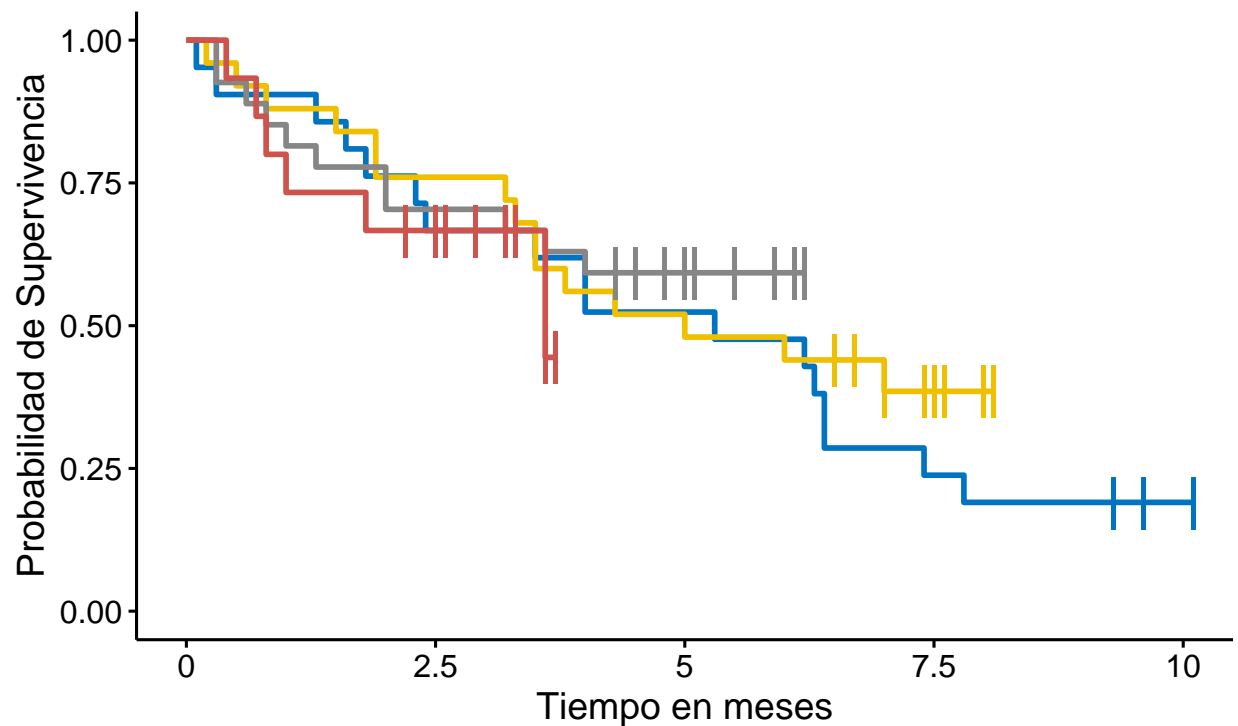
```
##
## Chisq= 11.1 on 8 degrees of freedom, p= 0.2
survdifff(Surv(time,delta)~diagyr,rho=1,data=larynx) # Generalized Wilcoxon

## Call:
## survdiff(formula = Surv(time, delta) ~ diagyr, data = larynx,
##          rho = 1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## diagyr=70  2    0.409    1.32    0.6258    0.9712
## diagyr=71 12    5.335    5.69    0.0220    0.0363
## diagyr=72  9    5.524    3.71    0.8819    1.3604
## diagyr=73 11    2.945    6.27    1.7665    3.1081
## diagyr=74 14    7.624    5.05    1.3079    2.0008
## diagyr=75  8    2.104    3.46    0.5312    0.7666
## diagyr=76 19    6.844    6.40    0.0308    0.0488
## diagyr=77 11    5.060    2.80    1.8167    2.4044
## diagyr=78  4    0.000    1.13    1.1344    1.3912
##
## Chisq= 10.8 on 8 degrees of freedom, p= 0.2
```

Rechazamos, por lo que no hay una diferencia, es decir, el año de diagnóstico parece no afectar con este agrupamiento Agrupemos...

```
aux5<- larynx %>%
  mutate("Fecha_Diag"=cut(larynx$diagyr,breaks=seq(70,78,by=2)))
year_fit2 <- survfit(Surv_general ~ Fecha_Diag, type = "kaplan-meier",
                    conf.int = 0.95,conf.type = "plain",data=aux5)
ggsurvplot(year_fit2,data = aux5,palette = pal_jco()(9),censor.size=7,
            censor.shape= 124,conf.int = FALSE,
            xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```

Strata + Fecha_Diag=(70,72] + Fecha_Diag=(72,74] + Fecha_Diag=(74,76] + Fecha_Diag=(76,78]



Así no parece haber mucha diferencia, podemos ver que se agrupan demasiado en los primeros 4 meses, hagamos el test:

```
print("Para el segundo grupo de años donde los grupos son: ")
```

```
## [1] "Para el segundo grupo de años donde los grupos son: "
```

```
print(levels(aux5$Fecha_Diag))
```

```
## [1] "(70,72]" "(72,74]" "(74,76]" "(76,78]"
```

```
survdif(Surv(time,delta)~Fecha_Diag,rho=0,data=aux5) # log-rank
```

```
## Call:
```

```
## survdif(formula = Surv(time, delta) ~ Fecha_Diag, data = aux5,
```

```
##      rho = 0)
```

```
##
```

```
## n=88, 2 observations deleted due to missingness.
```

```
##
```

```
##
```

```
## Fecha_Diag=(70,72] 21      17      14.25      0.532      0.790
```

```
## Fecha_Diag=(72,74] 25      15      16.81      0.194      0.310
```

```
## Fecha_Diag=(74,76] 27      11      13.17      0.359      0.536
```

```
## Fecha_Diag=(76,78] 15       6       4.77      0.316      0.377
```

```
##
```

```
## Chisq= 1.4 on 3 degrees of freedom, p= 0.7
```

```
survdif(Surv(time,delta)~Fecha_Diag,rho=1,data=aux5) # Generalized Wilcoxon
```

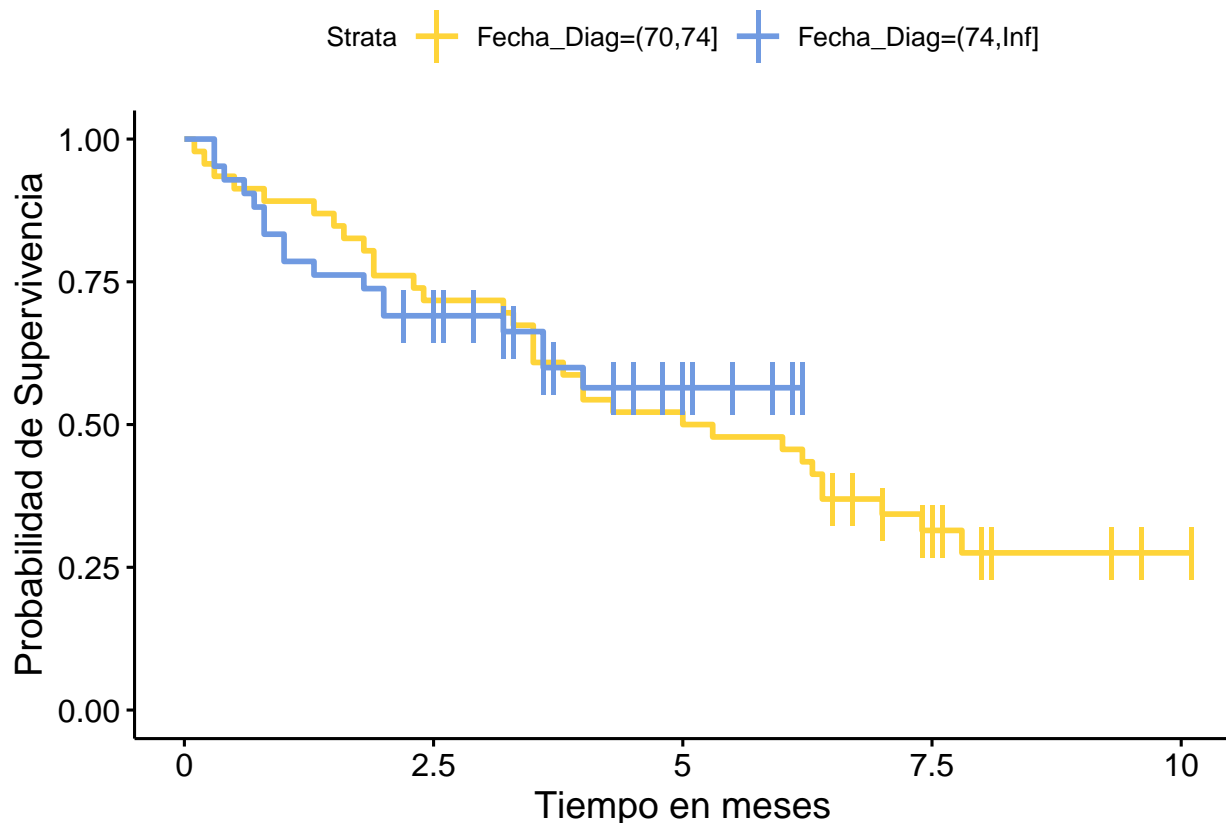
```
## Call:
```

```
## survdiff(formula = Surv(time, delta) ~ Fecha_Diag, data = aux5,
##      rho = 1)
##
## n=88, 2 observations deleted due to missingness.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Fecha_Diag=(70,72] 21    10.73    9.58    0.1394    0.266
## Fecha_Diag=(72,74] 25    10.48   11.44    0.0809    0.165
## Fecha_Diag=(74,76] 27     8.90   10.10    0.1433    0.268
## Fecha_Diag=(76,78] 15     5.04    4.03    0.2531    0.355
##
## Chisq= 0.8  on 3 degrees of freedom, p= 0.9
```

Rechazamos, por lo que no hay una diferencia, es decir, el año de diagnóstico parece no afectar con este agrupamiento

Probemos con otro agrupamiento:

```
aux6<- larynx %>%
  mutate("Fecha_Diag"=cut(larynx$diagyr,breaks=c(70,74,Inf)))
year_fit3 <- survfit(Surv_general ~ Fecha_Diag, type = "kaplan-meier",
  conf.int = 0.95, conf.type = "plain",data=aux6)
ggsurvplot(year_fit3,data = aux6,palette = pal_simpsons()(9),censor.size=7,
  censor.shape= 124,conf.int = FALSE,
  xlab = "Tiempo en meses", ylab = "Probabilidad de Supervivencia")
```



De nuevo, aquí no parece haber ninguna diferencia. Hagamos los test:

```

print("Para el tercer grupo de años donde los grupos son: ")

## [1] "Para el tercer grupo de años donde los grupos son: "
print(levels(aux6$Fecha_Diag))

## [1] "(70,74]" "(74,Inf]"

survdifff(Surv(time,delta)~Fecha_Diag,rho=0,data=aux6) # log-rank

## Call:
## survdifff(formula = Surv(time, delta) ~ Fecha_Diag, data = aux6,
##           rho = 0)
##
## n=88, 2 observations deleted due to missingness.
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## Fecha_Diag=(70,74] 46      32      31.1    0.0289    0.0907
## Fecha_Diag=(74,Inf] 42      17      17.9    0.0500    0.0907
##
## Chisq= 0.1  on 1 degrees of freedom, p= 0.8

survdifff(Surv(time,delta)~Fecha_Diag,rho=1,data=aux6) # Generalized Wilcoxon

## Call:
## survdifff(formula = Surv(time, delta) ~ Fecha_Diag, data = aux6,
##           rho = 1)
##
## n=88, 2 observations deleted due to missingness.
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## Fecha_Diag=(70,74] 46      21.2     21.0   0.00178   0.00606
## Fecha_Diag=(74,Inf] 42      13.9     14.1   0.00265   0.00606
##
## Chisq= 0  on 1 degrees of freedom, p= 0.9

```

Rechazamos, por lo que no hay una diferencia, es decir, el año de diagnóstico parece no afectar con este agrupamiento

En conclusión: El año de diagnóstico no parece ser una variable importante, bajo ningún agrupamiento, mientras que la etapa sí lo es, y la edad del paciente bajo el agrupamiento de 40 a 70 y de 71 en adelante. Cabe destacar que obtuvimos mejores resultados agrupando la etapa 1 y 2 en una sola, y aún mejores resultados agrupando la 1 y 2 en una sola y la 3 y 4 en otra distinta, obteniendo la clasificación “Temprana” y “Avanzada”. # Pruebas Log-Rank con $\alpha = 0.05$

Ejercicio 3

Usando pruebas No paramétricas Log-Rank:

Por estadio

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

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

```

survdifff(Surv(time,delta)~stage,rho=0,data=larynx) #log-rank

```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage, data = larynx,
##          rho = 0)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## stage=1 33         15    22.57      2.537      4.741
## stage=2 17          7    10.01      0.906      1.152
## stage=3 27         17    14.08      0.603      0.856
## stage=4 13         11     3.34     17.590     19.827
##
##  Chisq= 22.8  on 3 degrees of freedom, p= 5e-05
```

```
survdiff(Surv(time,delta)~stage,rho=1,data=larynx) # Generalized Wilcoxon
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage, data = larynx,
##          rho = 1)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## stage=1 33      9.26    15.78      2.694      6.48
## stage=2 17      4.66     7.18      0.884      1.47
## stage=3 27     12.77    10.06      0.730      1.35
## stage=4 13      9.15     2.82     14.219     18.78
##
##  Chisq= 23.1  on 3 degrees of freedom, p= 4e-05
```

Por lo que, con ambos test obtenemos un p-value demasiado pequeño, entonces podemos decir que hay evidencia estadísticamente significativa como para decir que son funciones de supervivencia distintas; veamos con la agrupación que hicimos:

```
survdiff(Surv(time,delta)~stage_aux,rho=0,data=stage_aux) #log-rank
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage_aux, data = stage_aux,
##          rho = 0)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## stage_aux=1 o 2 50      22    32.58      3.435     10.131
## stage_aux=3      27      17    14.08      0.603      0.856
## stage_aux=4      13      11     3.34     17.590     19.827
##
##  Chisq= 22.7  on 2 degrees of freedom, p= 1e-05
```

```
survdiff(Surv(time,delta)~stage_aux,rho=1,data=stage_aux) # Generalized Wilcoxon
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage_aux, data = stage_aux,
##          rho = 1)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## stage_aux=1 o 2 50     13.93    22.97      3.56     13.14
## stage_aux=3      27     12.77    10.06      0.73      1.35
## stage_aux=4      13      9.15     2.82     14.22     18.78
##
##  Chisq= 23.1  on 2 degrees of freedom, p= 1e-05
```

¡Tenemos un p-value muy pequeño! Por lo que sí hay una diferencia significativa.


```
survdif(Surv(time,delta)~stage_aux,rho=0,data=stage_aux2) #log-rank
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage_aux, data = stage_aux2,
##      rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## stage_aux=Avanzada(3,4) 40      28    17.4      6.42     10.1
## stage_aux=Temprana(1,2) 50      22    32.6      3.43     10.1
##
##  Chisq= 10.1  on 1 degrees of freedom, p= 0.001
```

```
survdif(Surv(time,delta)~stage_aux,rho=1,data=stage_aux2) # Generalized Wilcoxon
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ stage_aux, data = stage_aux2,
##      rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## stage_aux=Avanzada(3,4) 40      21.9    12.9      6.35     13.1
## stage_aux=Temprana(1,2) 50      13.9    23.0      3.56     13.1
##
##  Chisq= 13.1  on 1 degrees of freedom, p= 3e-04
```

Bajo la clasificación de “Temprana” y “Avanzada”, el p-value es aún menor ## Por grupo de edad Compare las funciones de supervivencia por grupos de edad

```
print("Para el primer grupo de edad donde los grupos son: ")
```

```
## [1] "Para el primer grupo de edad donde los grupos son: "
print(levels(aux$Grupo))
```

```
## [1] "(40,55]" "(55,70]" "(70,Inf]"
```

```
survdif(Surv(time,delta)~Grupo,rho=0,data=aux) # log-rank
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ Grupo, data = aux, rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,55] 21        9     10.8     0.300     0.392
## Grupo=(55,70] 41       21     25.9     0.913     1.941
## Grupo=(70,Inf] 28       20     13.3     3.325     4.615
##
##  Chisq= 4.6  on 2 degrees of freedom, p= 0.1
```

```
survdif(Surv(time,delta)~Grupo,rho=1,data=aux) # Generalized Wilcoxon
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ Grupo, data = aux, rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,55] 21       7.04     8.02     0.12     0.202
## Grupo=(55,70] 41     13.94    18.09     0.95     2.562
## Grupo=(70,Inf] 28     14.86     9.74     2.70     4.885
##
##  Chisq= 5  on 2 degrees of freedom, p= 0.08
```

```

print("Para el segundo grupo de edad donde los grupos son: ")

## [1] "Para el segundo grupo de edad donde los grupos son: "
print(levels(aux2$Grupo))

## [1] "(40,65]" "(65,Inf]"

survdifff(Surv(time,delta)~Grupo,rho=0,data=aux2) # log-rank

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux2, rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,65] 46         23      24.9      0.147      0.298
## Grupo=(65,Inf] 44         27      25.1      0.146      0.298
##
## Chisq= 0.3  on 1 degrees of freedom, p= 0.6

survdifff(Surv(time,delta)~Grupo,rho=1,data=aux2) # Generalized Wilcoxon

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux2, rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,65] 46         17.2     18.1      0.0429     0.115
## Grupo=(65,Inf] 44         18.7     17.8      0.0436     0.115
##
## Chisq= 0.1  on 1 degrees of freedom, p= 0.7

print("Para el tercer grupo de edad donde los grupos son: ")

## [1] "Para el tercer grupo de edad donde los grupos son: "
print(levels(aux3$Grupo))

## [1] "(40,70]" "(70,Inf]"

survdifff(Surv(time,delta)~Grupo,rho=0,data=aux3) # log-rank

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux3, rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,70] 62         30      36.7      1.21      4.62
## Grupo=(70,Inf] 28         20      13.3      3.33      4.62
##
## Chisq= 4.6  on 1 degrees of freedom, p= 0.03

survdifff(Surv(time,delta)~Grupo,rho=1,data=aux3) # Generalized Wilcoxon

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux3, rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,70] 62         21.0     26.11      1.01      4.89
## Grupo=(70,Inf] 28         14.9      9.74      2.70      4.89
##
## Chisq= 4.9  on 1 degrees of freedom, p= 0.03

```

```

print("Para el cuarto grupo de edad donde los grupos son: ")

## [1] "Para el cuarto grupo de edad donde los grupos son: "
print(levels(aux4$Grupo))

## [1] "(40,50]" "(50,60]" "(60,70]" "(70,Inf]"

survdifff(Surv(time,delta)~Grupo,rho=0,data=aux4) # log-rank

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux4, rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,50] 11         5      5.3      0.017   0.0195
## Grupo=(50,60] 19         9     10.6     0.234   0.3023
## Grupo=(60,70] 32        16     20.8     1.103   1.9375
## Grupo=(70,Inf] 28        20     13.3     3.325   4.6151
##
##  Chisq= 4.8  on 3 degrees of freedom, p= 0.2

survdifff(Surv(time,delta)~Grupo,rho=1,data=aux4) # Generalized Wilcoxon

## Call:
## survdifff(formula = Surv(time, delta) ~ Grupo, data = aux4, rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=(40,50] 11      4.11      4.00  0.00304  0.00444
## Grupo=(50,60] 19      6.64      7.61  0.12160  0.20428
## Grupo=(60,70] 32     10.23     14.51  1.25967  2.84982
## Grupo=(70,Inf] 28     14.86      9.74  2.69861  4.88507
##
##  Chisq= 5.4  on 3 degrees of freedom, p= 0.1

```

Entonces, para el acomodo que hicimos partiendo en el grupo de edad de 40 a 70 y de 71 en adelante, ambos test nos dicen que sí son diferentes las funciones de supervivencia; para las demás clasificaciones nos dice que no hay evidencia significativa estadísticamente como para decir que sí son diferentes.