

Proyecto 1: Tarea 3

Alexis Adrian Carrillo Medina

Setup

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
library(survival)
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 4.1.1
```

```
## Loading required package: ggpubr
```

```
Sujeto <- 1:20
Edad <- c(37,40,42,44,46,48,49,50,51,52,52,53,56,57,59,62,63,63,64,65)
Ab_de_alc <- c(1,0,0,1,0,0,1,1,1,0,0,0,1,0,0,1,1,1,1,1)
remision_time <- c(0.3,1.8,1.7,1.5,2.5,3.7,2.9,2.3,2.8,4.6,3.8,5.3,3.6,2.6,2.9,5.5,2.8,4.0,6.0,4.3)
Censura <- c(1,1,1,1,1,1,1,0,1,1,1,1,1,1,0,1,0,1,1,1)

Data <- data.frame(Sujeto,Edad,Ab_de_alc,remision_time,Censura)
```

Ejercicio

Veinte sujetos son reclutados para un estudio clínico de un nuevo tratamiento para la cirrosis del hígado, una enfermedad crónica caracterizada por la pérdida de células hepáticas funcionales. Se cree que el nuevo tratamiento puede poner la cirrosis en remisión, si la enfermedad es causada por el abuso del alcohol. (La remisión es la atenuación o desaparición completa en el paciente de los signos y síntomas de su enfermedad, ya sea como consecuencia del tratamiento o de forma espontánea.)

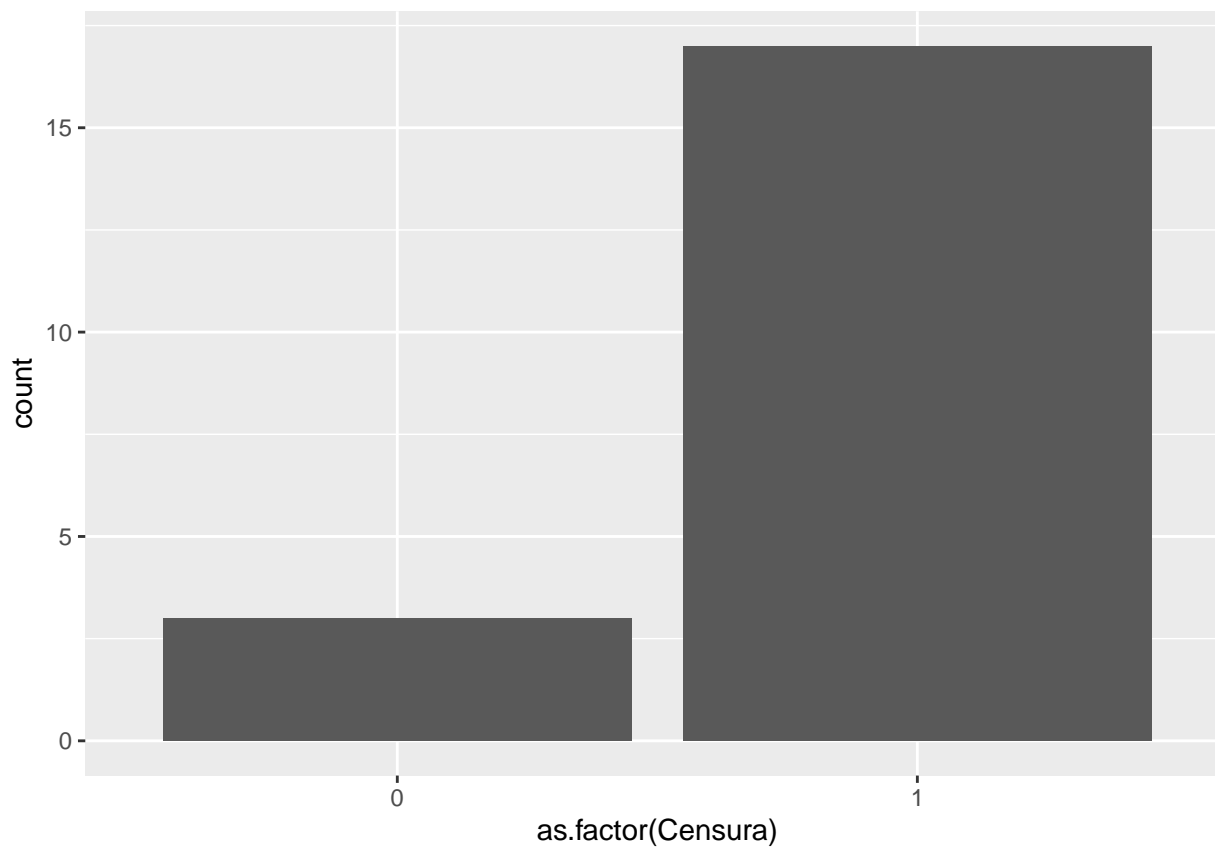
Las covariables son la edad de un sujeto (en años) y el indicador de abuso actual de alcohol ($s_1 = 1$, no = 0). El tiempo de supervivencia es el tiempo hasta una remisión (en semanas). Si un sujeto murió durante el ensayo se considera censurado. Si un sujeto estaba a la espera de la remisión, y continúa hasta al final del estudio, también se considera censurado.

Realice un análisis descriptivo sobre el tiempo de supervivencia de los sujetos y de sus características.

Contamos con 5 variables (Sujeto, Edad, Abuso de alcohol, Tiempo de remisión, Censura); Realizaremos gráficas de cómo se relaciona el tiempo de supervivencia de estos individuos con sus características

Primero observemos la variable de Censura, para saber como se comportaron los individuos bajo este experimento y saber si es significativo el experimento.

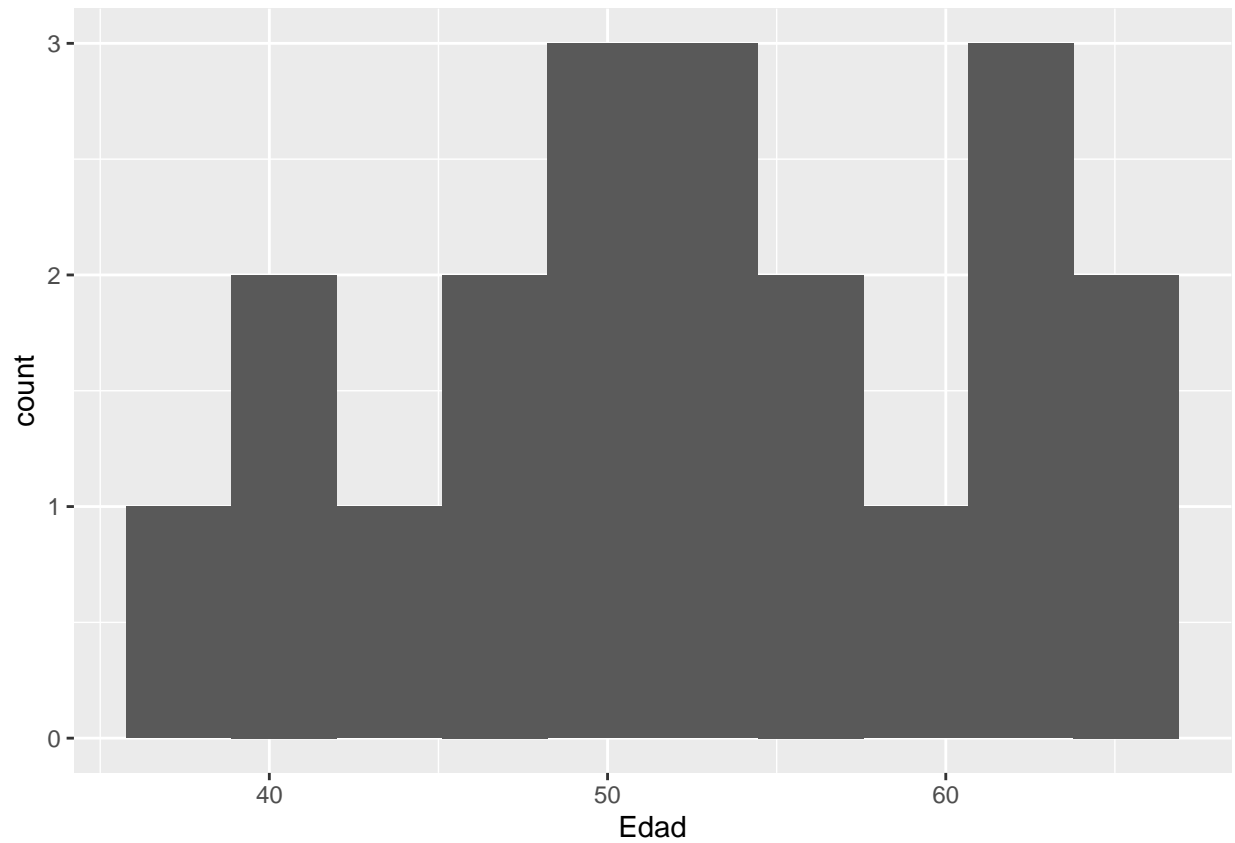
```
ggplot(data=Data, aes(x=as.factor(Censura)))+  
  geom_bar()
```



Lo que nos indica que la mayoría llegaron al tiempo de remision y entonces es significativo.

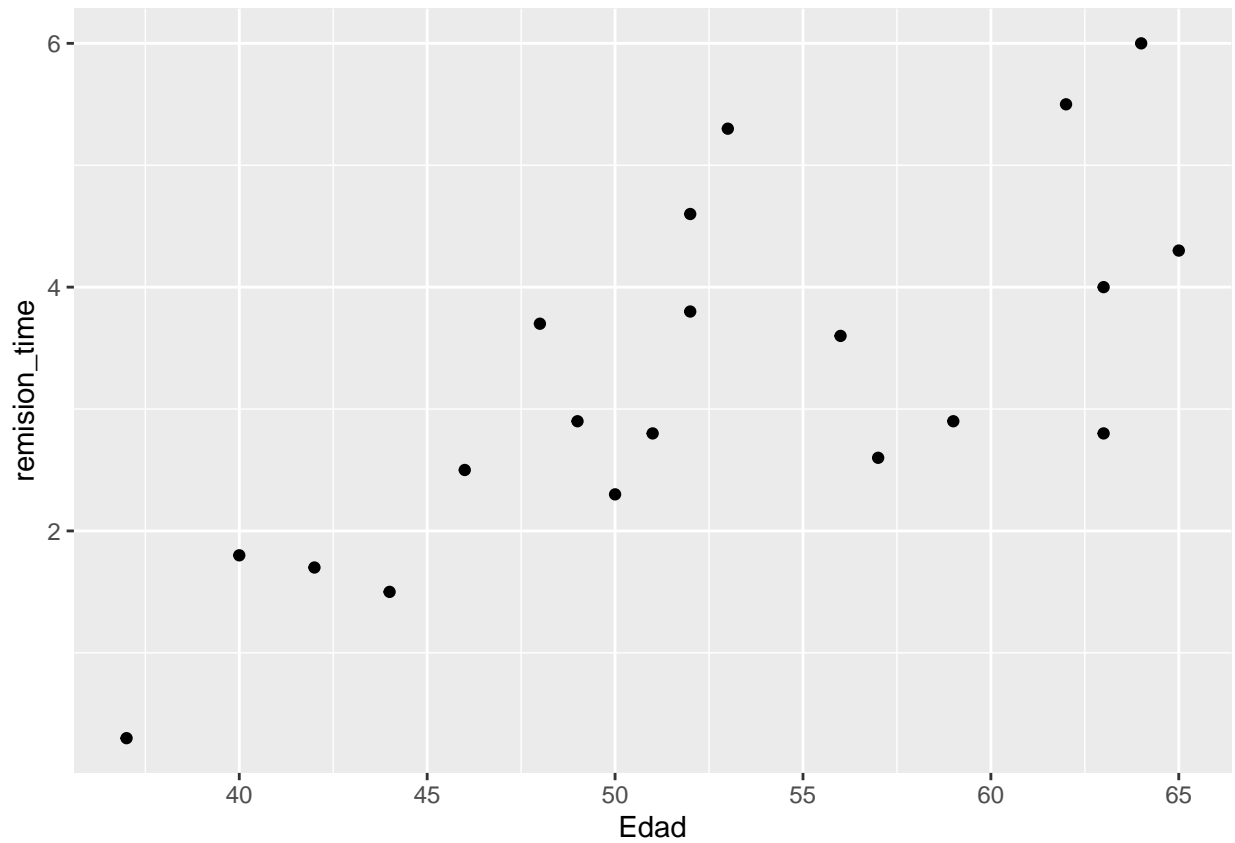
Veamos las variables ahora. Para la Edad, primero veamos un histograma

```
ggplot(data=Data, aes(Edad)) +  
  geom_histogram(bins=10)
```



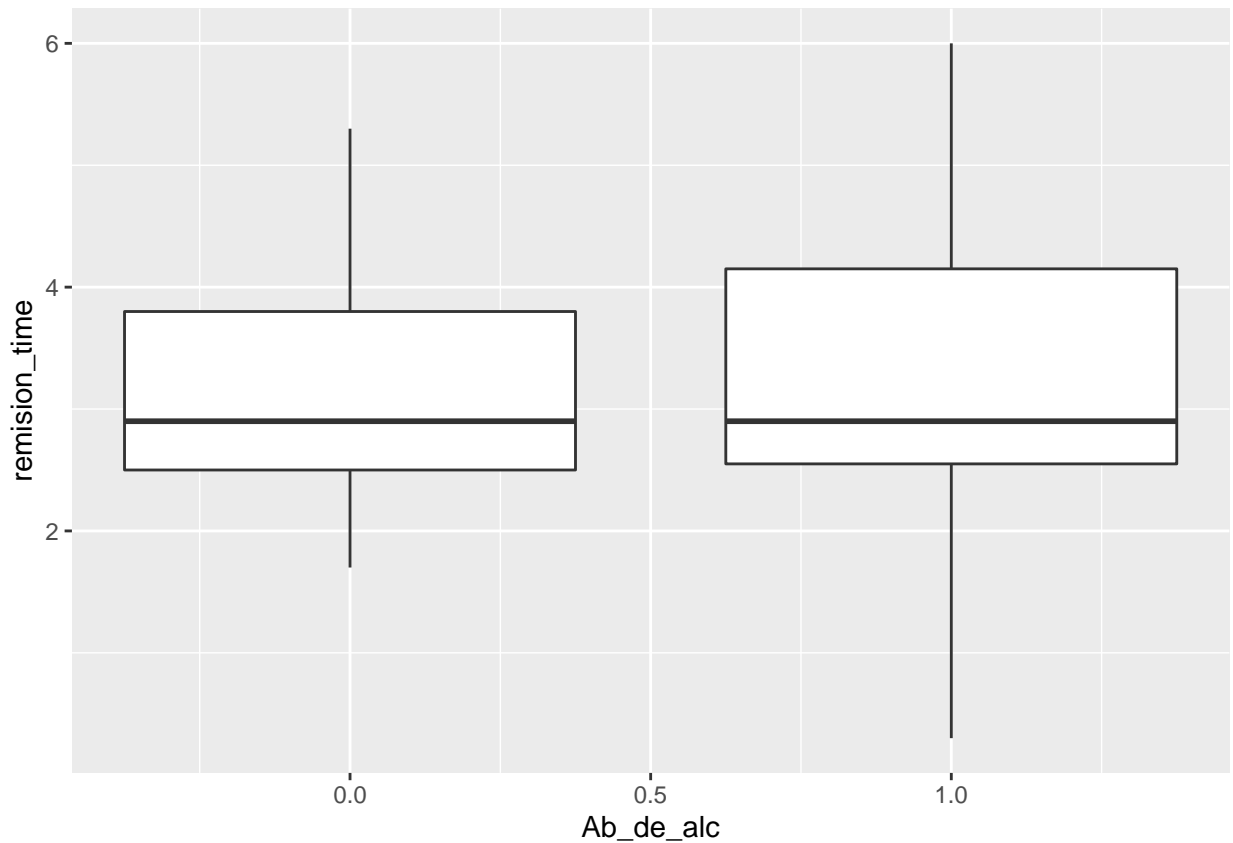
Observemos que la mayoría de la genta es de edad 40+. Veamos como se relaciona esto con el tiempo de remision

```
ggplot(data=Data,aes(x=Edad,y=remision_time))+  
  geom_point()
```



Donde notamos que existe una leve relacion lineal entre la edad y el tiempo de remision: Entre mas edad mas tiempo de remision. Ahora veamos para abuso de alcohol

```
ggplot(data=Data,aes(group=Ab_de_alc,x=Ab_de_alc,y=remision_time))+
  geom_boxplot()
```



De aqui notamos que el grupo que abusa del alchol y el grupo que no, no tienen grandes diferencias en el tiempo de remision, mas que el tercer cuartil del grupo que abuso del alcohol es ligeramente mayor.

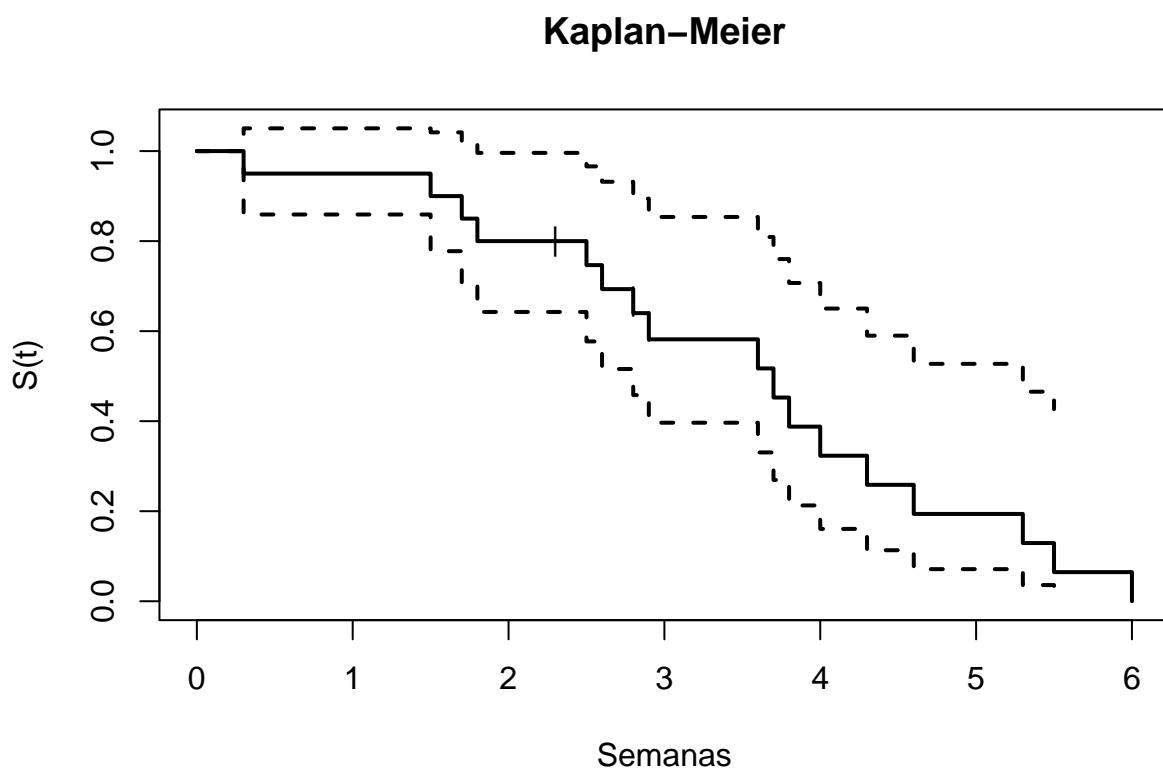
Estime la funcion de Supervivencia $S(t)$ usando el metodo de Kaplan-Meier. Grafique

$S(t)$ junto con sus intervalos de confianza del 95% para:

- La poblacion completa.

Veamos primero el modelo para la poblacion completa, utilizando Kaplan-Meier

```
model1 <- survfit(Surv(remision_time,Censura)~1, type = "kaplan-meier", conf.type="none")
plot(model1,mark="|",lwd=2,xlab="Semanas",ylab="S(t)",main="Kaplan-Meier")
```

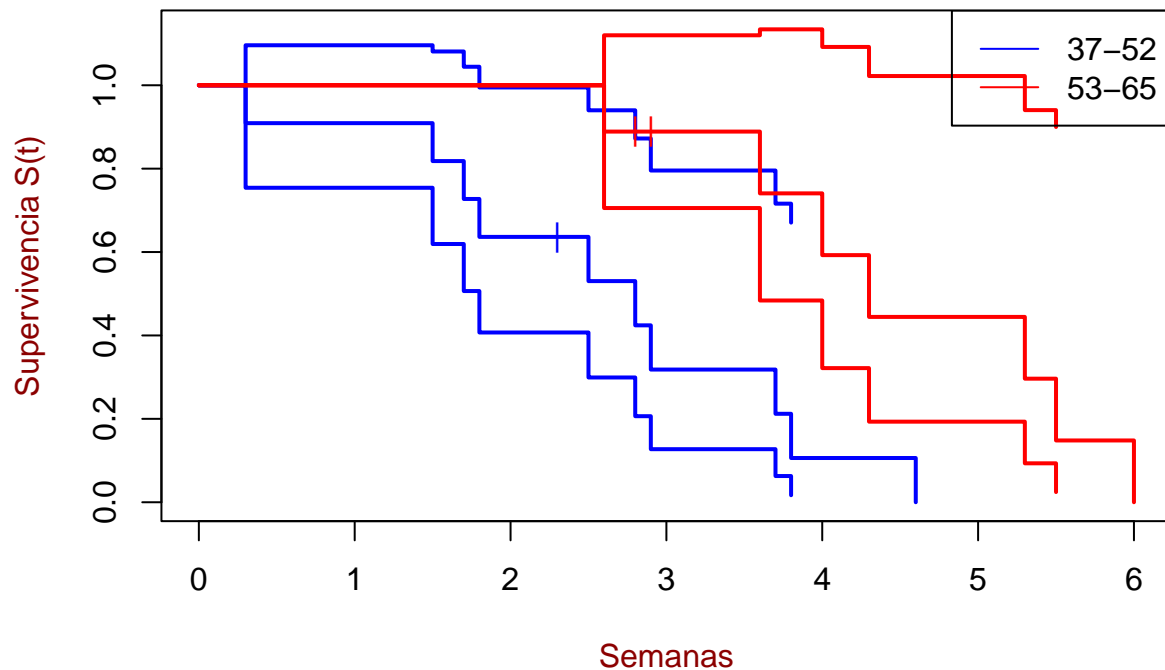


- Por grupos de edad: 37-52 y 53-65.

Agrupemos por poblacion

```
grupo = factor(Edad > 52)
model2 <- survfit(Surv(remision_time,Censura)~grupo, type = "kaplan-meier", conf.type="none")
plot(model2,mark="|",conf.int=T,lwd=2,xlab="Semanas", ylab="Supervivencia S(t)",lty=c(1,1),col=c("blue"
legend("topright", c("37-52", "53-65"), lty=c(1,1),col=c("blue","red"))
```

Comparacion de Edades



Inicialmente, pareciera que existen diferencias entre ambas funciones de supervivencia, no obstante los intervalos se entrecruzan, lo que indica que podria suceder que en realidad no existe un diferencia entre ambos grupos de edades, para ellos recurriremos al siguiente test estadistico

```
survdif(Surv(remision_time,Censura)~grupo,rho=0)
```

```
## Call:
## survdiff(formula = Surv(remision_time, Censura) ~ grupo, rho = 0)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=FALSE 11      10     5.16      4.53      7.59
## grupo=TRUE   9       7    11.84      1.98      7.59
##
## Chisq= 7.6  on 1 degrees of freedom, p= 0.006
```

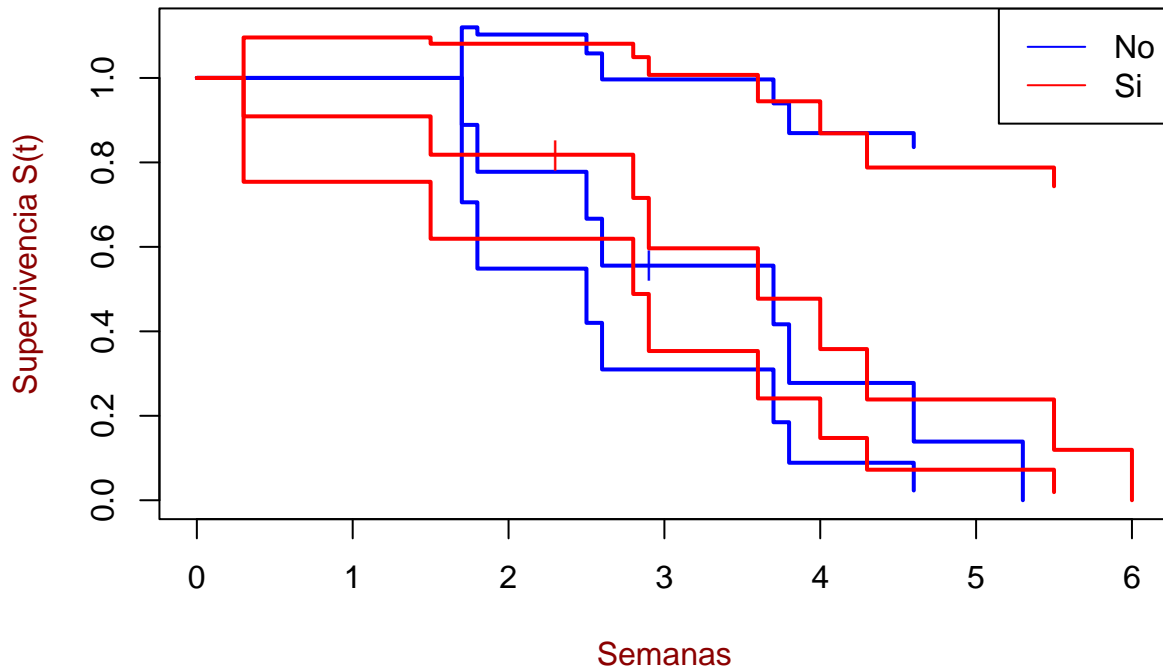
Como $p = 0.006 < 0.05$ entonces tenemos que si existe diferencia entre los 2 grupos.

- Por abuso de alcohol.

Realicemos el modelo por abuso

```
model3 <- survfit(Surv(remision_time,Censura)~as.factor(Ab_de_alc), type = "kaplan-meier", conf.type="n")
plot(model3,mark="|",conf.int=T,lwd=2,xlab="Semanas", ylab="Supervivencia S(t)",lty=c(1,1),col=c("blue",
legend("topright", c('No', "Si"), lty=c(1,1),col=c("blue","red"))
```

Comparacion por Abuso de alcohol



Entonces, a primera instancia, las curvas principales parecen tener diferencias, no obstante, se sobreponen una de otras, habra que volver a utilizar un test estadístico.

```
survdif(Surv(remision_time,Censura)~as.factor(Ab_de_alc),rho=0)
```

```
## Call:
## survdiff(formula = Surv(remision_time, Censura) ~ as.factor(Ab_de_alc),
##          rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(Ab_de_alc)=0  9         8      6.57    0.312    0.561
## as.factor(Ab_de_alc)=1 11         9     10.43    0.196    0.561
##
##  Chisq= 0.6  on 1 degrees of freedom, p= 0.5
```

Entonces, como $p = 0.5$ tenemos que como tal no podemos rechazar la hipótesis de que no existan diferencias.

Ajuste un modelo de riesgos proporcionales de Cox para definir la contribucion de las

variables al tiempo de supervivencia de los pacientes.

Probemos un modelo donde consideremos solo consideremos el abuso de alcohol y a la edad (Puesto que vimos una tendencia lineal en la edad.)


```
cox1 <- coxph(Surv(remision_time,Censura)~Ab_de_alc+Edad)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(remision_time, Censura) ~ Ab_de_alc + Edad)
##
##   n= 20, number of events= 17
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Ab_de_alc  1.81963   6.16960  0.81158  2.242 0.024956 *
## Edad      -0.30958   0.73375  0.08263 -3.746 0.000179 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Ab_de_alc    6.1696    0.1621    1.257   30.2744
## Edad         0.7338    1.3629    0.624    0.8628
##
## Concordance= 0.868 (se = 0.051 )
## Likelihood ratio test= 23.45 on 2 df,  p=8e-06
## Wald test              = 14.17 on 2 df,  p=8e-04
## Score (logrank) test = 23.13 on 2 df,  p=9e-06
```

Parece ser que la variable de Abuso de Alcohol no es significativa, veamos un modelo sin ella.

```
cox1 <- coxph(Surv(remision_time,Censura)~Edad)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(remision_time, Censura) ~ Edad)
##
##   n= 20, number of events= 17
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Edad -0.20675    0.81323  0.06015 -3.437 0.000587 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Edad    0.8132    1.23    0.7228    0.915
##
## Concordance= 0.842 (se = 0.05 )
## Likelihood ratio test= 17.49 on 1 df,  p=3e-05
## Wald test              = 11.82 on 1 df,  p=6e-04
## Score (logrank) test = 15.92 on 1 df,  p=7e-05
```

Observamos una mejora en el modelo, entonces este sera el adecuado.

-¿Cual es la estimacion puntual para los coeficientes de regresion? Interprete los coeficientes de regresion.

Veamos la estimacion

```
coef(cox1)
```

```
##      Edad  
## -0.2067464
```

Los coeficientes del modelo de cox se refieren a la relacion que hay de supervivencia, en este caso como son el logaritmo de un radio, entonces un coeficiente negativo, como este, significa que entre mas edad mas es la supervivencia, que en este caso significa mayor tiempo de remision.

-Obtenga un intervalo de confianza al 95% para la estimacion de los coeficientes de regresion.

```
confint(cox1)
```

```
##      2.5 %      97.5 %  
## Edad -0.3246287 -0.0888641
```

Entonces el intervalo queda

$$\beta_{inf} = -0.3246287, \hat{\beta}_1 = 0.2067464, \beta_{sup} = -0.0888641$$

-¿Las variables explicativas tienen o no efecto en el modelo? Justifique.

Si, como vemos en el siguiente codigo

```
summary(cox1)
```

```
## Call:  
## coxph(formula = Surv(remision_time, Censura) ~ Edad)  
##  
##      n= 20, number of events= 17  
##  
##      coef exp(coef) se(coef)      z Pr(>|z|)  
## Edad -0.20675    0.81323   0.06015 -3.437 0.000587 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
##      exp(coef) exp(-coef) lower .95 upper .95  
## Edad    0.8132      1.23    0.7228    0.915  
##  
## Concordance= 0.842 (se = 0.05 )  
## Likelihood ratio test= 17.49 on 1 df,  p=3e-05  
## Wald test            = 11.82 on 1 df,  p=6e-04  
## Score (logrank) test = 15.92 on 1 df,  p=7e-05
```

La variable edad tiene un $p - value = 0.0005$ lo cual indica que es significativo

-¿Cual es la proporcion del riesgo de falla del abuso de alcohol?

Ajustemos un modelo con abuso de alcohol

```
cox1 <- coxph(Surv(remision_time,Censura)~Ab_de_alc+Edad)  
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(remision_time, Censura) ~ Ab_de_alc + Edad)
##
## n= 20, number of events= 17
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Ab_de_alc  1.81963   6.16960  0.81158  2.242 0.024956 *
## Edad      -0.30958   0.73375  0.08263 -3.746 0.000179 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Ab_de_alc    6.1696    0.1621    1.257   30.2744
## Edad         0.7338    1.3629    0.624    0.8628
##
## Concordance= 0.868 (se = 0.051 )
## Likelihood ratio test= 23.45 on 2 df,  p=8e-06
## Wald test            = 14.17 on 2 df,  p=8e-04
## Score (logrank) test = 23.13 on 2 df,  p=9e-06
```

Recordemos que los coeficientes del modelo de cox no son mas que el logaritmo del ratio entre 2 individuos que se diferencian entre la variable explicativa. Entonces tomando la exponencial de este coeficiente podemos ver en proporcion como cambia el riesgo de falla. Usando el codigo anterior tenemos que la proporcion por riesgo de falla por abuso de alcohol es 6.1696

-¿Los sujetos que tienen un abuso de alcohol tienen mayor riesgo de entrar en remision?

Dado que la proporcion anterior es positiva, nos indicaria que si. Aquellos que abusaron del alcohol tienen mayor riesgo de entrar en remision.