



Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas



Unidad de aprendizaje: Estadística Aplicada

PIA.- Análisis de Supervivencia

Marco Antonio Oviedo Acevedo 1851698

Camila Cruz Pineda 1911855

Fernanda Montserrat Chávez Rivera 1841926

Maestro: MET. Alejandra Cerda

Monterrey, 17 de noviembre del 2022

Análisis de Supervivencia (transplant)

17 de noviembre del 2022

Contents

ANÁLISIS DE SUPERVIVENCIA	2
1. DESCRIPCIÓN DE LOS DATOS	2
2. OBJETIVO DE LA INVESTIGACIÓN	2
2.1 GENERAL	2
2.2 ESPECÍFICOS	3
2.3 ALCANCE DEL PROYECTO	3
3. ANÁLISIS DE LOS DATOS	3
3.1 PREPARACION DE LOS DATOS	3
3.1.1 CARGA DE LOS DATOS [ITEMS PARES]	3
3.1.2 LIMPIEZA VALORES DUPLICADOS	3
3.1.3 LIMPIEZA VALORES NULOS	4
3.1.3 AJUSTE DE LOS VALORES CENSURADOS	6
3.2 ANÁLISIS DESCRIPTIVO DE LOS DATOS	6
3.3 ANÁLISIS DE REGRESIÓN	11
3.4 ANÁLISIS DE SUPERVIVENCIA	13
3.4.1 MODELO COX	13
3.4.2 ANÁLISIS DE LOS RESIDUOS DEL MODELO COX	18
3.4.2 COMPARACIÓN DEL MODELO COX VS MODELO DE REGRESIÓN	21
3.4.3 MODELO KAPLAN-MEIER	21
3.4.4 MODELOS PARAMÉTRICOS	26
3.4.4 COMPARACIÓN DE LOS MODELOS PARAMETRICOS	33
4. CONCLUSIÓN	35
5. BIBLIOGRAFÍAS	36

ANÁLISIS DE SUPERVIVENCIA

1. DESCRIPCIÓN DE LOS DATOS

Nombre: transplant {survival} [items pares] Lista de espera para trasplante de hígado

Descripción: Sujetos en una lista de espera de trasplante de hígado de 1990-1999, y su disposición: recibieron un trasplante, murieron mientras esperaban, se retiraron de la lista o fueron censurados.

Uso (código): trasplante data(transplant, package="survival")

Formato: Un marco de datos con 815 observaciones (trasplante) sobre las siguientes 6 variables.

- *age:* edad al agregar a la lista de espera
- *sex:* m o f
- *abo:* tipo de sangre: A, B, AB u O
- *year:* año en que entraron en lista de espera
- *futime:* tiempo desde la entrada hasta la disposición final
- *event:* Disposición final: censurado, muerte, LTX o retirarse

Detalles: Esto representa la experiencia del trasplante en una región en particular, durante un período de tiempo en el que el trasplante de hígado se reconoció mucho más ampliamente como una modalidad de tratamiento viable. El número de trasplantes de hígado aumenta durante el período, pero el número de sujetos agregados a la lista de espera de trasplante de hígado creció mucho más rápido. Las preguntas importantes abordadas por los datos son el cambio en el tiempo de espera, quién espera y si hubo un aumento consecuente en las muertes mientras estaban en la lista.

El tipo de sangre es una consideración importante. Los hígados de donantes de sujetos con tipo de sangre O pueden ser utilizados por pacientes con tipos de sangre A, B, AB o O, mientras que un hígado AB solo puede ser utilizado por un receptor AB. Por lo tanto, los sujetos de tipo O en la lista de espera están en desventaja, ya que el grupo de competidores es mayor para los hígados de donantes de tipo O.

Estos datos son de interés histórico y proporcionan un ejemplo útil de riesgos competitivos, pero tienen poca relevancia para la práctica actual. Las políticas de asignación hepática han evolucionado y ahora dependen directamente del riesgo y la necesidad de cada paciente individual, cuyas evaluaciones se actualizan regularmente mientras el paciente está en la lista de espera. Sin embargo, la escasez general de órganos sigue siendo aguda.

Kim WR, Therneau TM, Benson JT, Kremers WK, Rosen CB, Gores GJ, Dickson ER. Muertes en la lista de espera de trasplante de hígado: un análisis de riesgos competitivos. Hepatología 2006 Feb; 43(2):345-51.

2. OBJETIVO DE LA INVESTIGACIÓN

2.1 GENERAL

Realizar un análisis de supervivencia para los sujetos en una lista de espera de trasplante de hígado de 1990-1999, y su disposición

2.2 ESPECÍFICOS

- Estimar la función de supervivencia a través del estimador de Kaplan y Meier para los sujetos en la lista de espera del trasplante de hígado de 1990-1999.
- Seleccionar el mejor modelo de regresión de riesgos proporcionales (Modelo de Cox), para obtener los mejores predictores de la supervivencia de los sujetos en la lista de espera del trasplante de hígado de 1990-1999.
- Interpretar los resultados de los mejores modelos de Cox obtenidos.
- Dar a comprender que características de los pacientes en lista de espera son los que necesitan mayor atención dada su probabilidad de supervivencia

2.3 ALCANCE DEL PROYECTO

Poder describir las características de los pacientes que tienen menor probabilidad de supervivencia en la lista de espera, esto para poder seleccionar a los pacientes que necesitan más los trasplantes.

3. ANÁLISIS DE LOS DATOS

3.1 PREPARACION DE LOS DATOS

3.1.1 CARGA DE LOS DATOS [ITEMS PARES]

Filtramos de mi base de datos “transplant” los items pares (2,4,6...)

```
##      age sex abo year futime      event
## 2    55  m   A 1991      28        ltx
## 4    40  f   0 1995     231        ltx
## 6    66  f   0 1996      58        ltx
## 8    55  m   0 1998      30        ltx
## 10   61  m   A 1991     139        ltx
## 12   61  m   A 1999     451 censored
```

3.1.2 LIMPIEZA VALORES DUPLICADOS

Verificamos si contamos con valores duplicados en las filas, esto para no generar errores en el análisis descriptivo y estadístico.

```
sum(duplicated(df)) # no tenemos valores duplicados
```

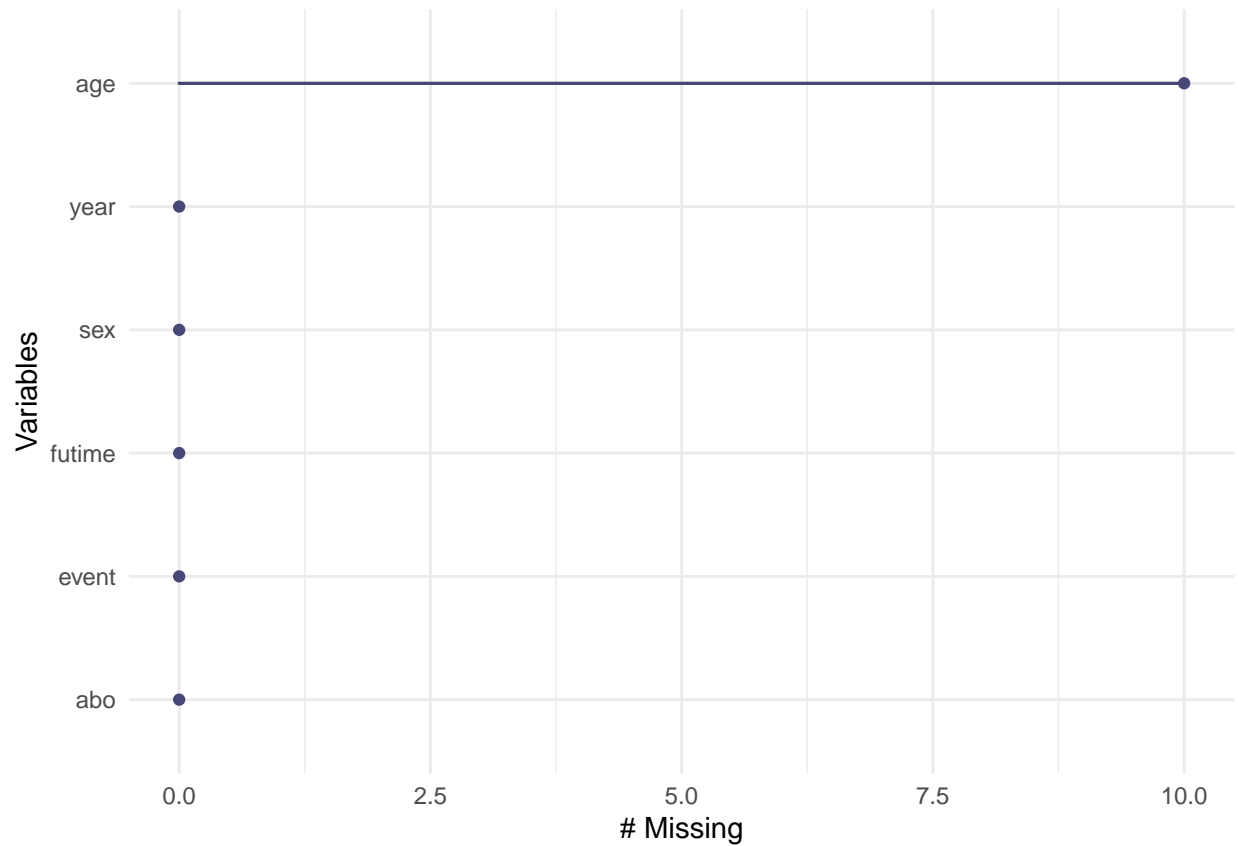
```
## [1] 0
```

Observamos que no contamos con filas duplicadas, entonces no tenemos que limpiar valores duplicados.

3.1.3 LIMPIEZA VALORES NULOS

Verificamos si contamos con valores nulos

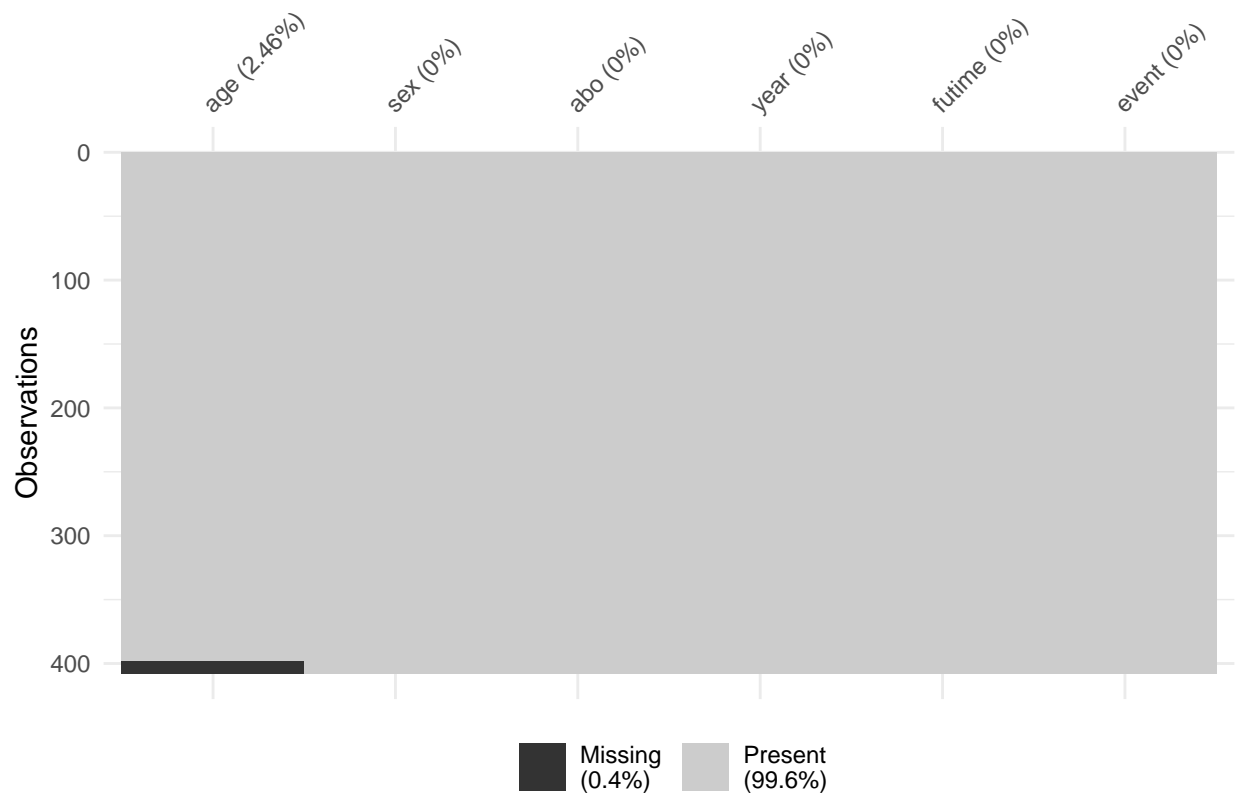
```
## [1] "La variable age tiene: 10 nulos"  
## [1] "La variable sex tiene: 0 nulos"  
## [1] "La variable abo tiene: 0 nulos"  
## [1] "La variable year tiene: 0 nulos"  
## [1] "La variable futime tiene: 0 nulos"  
## [1] "La variable event tiene: 0 nulos"
```



Observando el resultado así como la gráfica podemos observar que si contamos con valores nulos en la variable *age*, teniendo un total de 10 valores nulos

Ahora verificamos si contamos con más del 20% del valor nulo de la variable *age*:

```
## [1] "La variable age tiene: 2.46% nulos"  
## [1] "La variable sex tiene: 0% nulos"  
## [1] "La variable abo tiene: 0% nulos"  
## [1] "La variable year tiene: 0% nulos"  
## [1] "La variable futime tiene: 0% nulos"  
## [1] "La variable event tiene: 0% nulos"
```

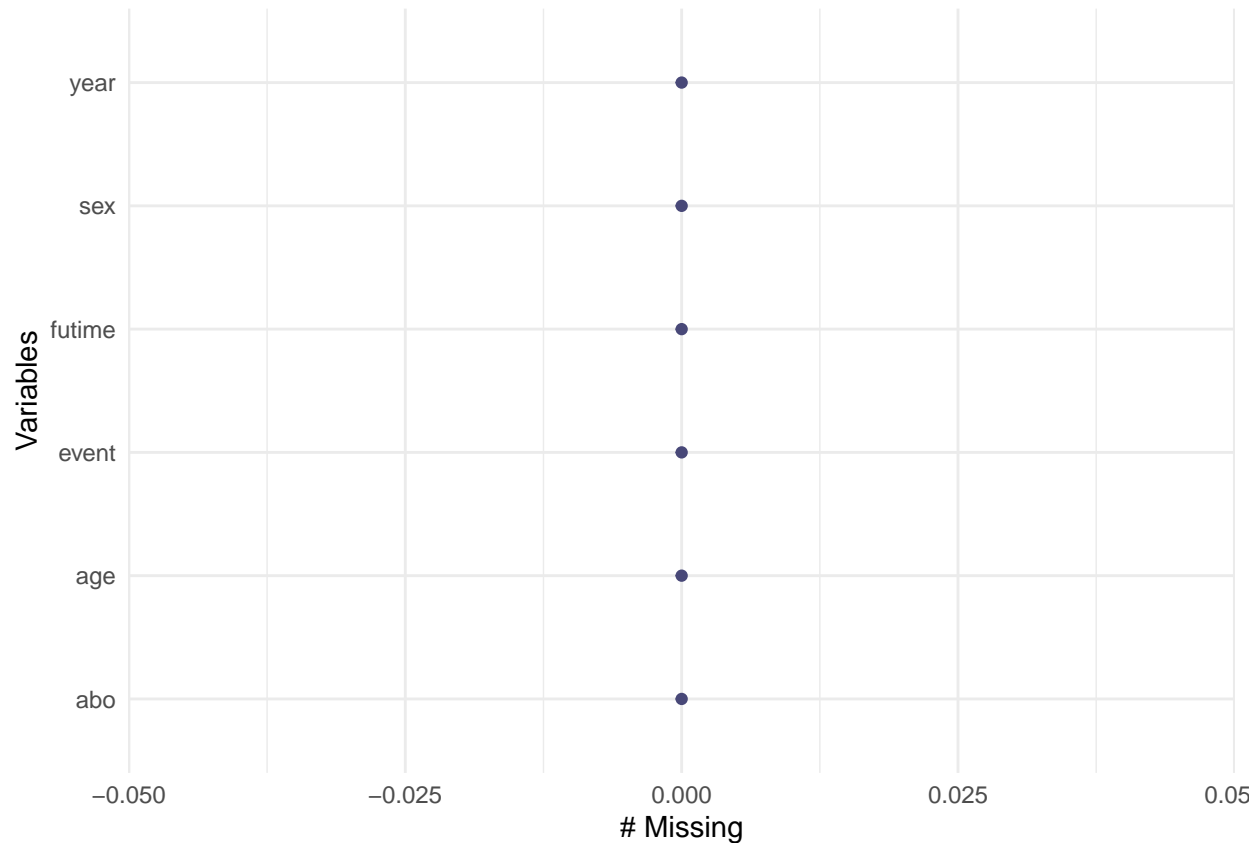


Dado que solo se cuenta con el 2.4% de valores nulos, podemos eliminar las filas con aquellos valores, pues no afectara al análisis

```
df <- na.omit(df)
```

verificamos que no tengamos valores nulos:

```
## [1] "La variable age tiene: 0 nulos"
## [1] "La variable sex tiene: 0 nulos"
## [1] "La variable abo tiene: 0 nulos"
## [1] "La variable year tiene: 0 nulos"
## [1] "La variable futime tiene: 0 nulos"
## [1] "La variable event tiene: 0 nulos"
```



la data ya no cuenta con valores nulos ni duplicados, entonces podemos trabajar con el análisis

3.1.3 AJUSTE DE LOS VALORES CENSURADOS

Creamos dos nuevas columnas para poder mejorar la sintaxis del estudio

```
df <- df %>%
  mutate(censura = if_else(event %in% c("ltx","censored","withdraw"),
                           0, 1)) %>%
  mutate(censura_name = if_else(event %in% c("ltx","censored","withdraw"),
                                "censuras", "muertes"))
```

3.2 ANÁLISIS DESCRIPTIVO DE LOS DATOS

Realizamos un resumen de mis datos:

```
summary(df) # Resumen de mis datos
```

##	age	sex	abo	year	future
##	Min. :18.00	m:210	A :154	Min. :1990	Min. : 0.0
##	1st Qu.:45.00	f:187	B : 50	1st Qu.:1993	1st Qu.: 52.0
##	Median :53.00		AB: 22	Median :1996	Median : 137.0
##	Mean :51.37		O :171	Mean :1995	Mean : 224.7
##	3rd Qu.:58.00			3rd Qu.:1998	3rd Qu.: 294.0

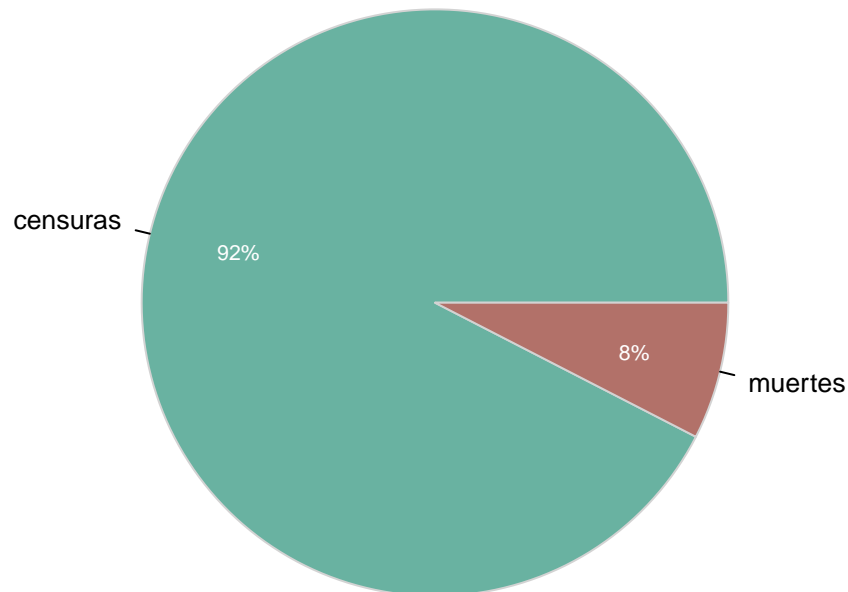
```
## Max.      :72.00                      Max.      :1999    Max.      :1899.0
##      event      censura      censura_name
## censored: 46    Min.      :0.00000    Length:397
## death      : 30    1st Qu.:0.00000    Class  :character
## ltx        :304    Median  :0.00000    Mode   :character
## withdraw: 17    Mean     :0.07557
##                      3rd Qu.:0.00000
##                      Max.      :1.00000
```

con el resumen previo de mis datos podemos notar lo siguiente:

- La edad promedio es de 51.37
- El tiempo promedio es de 224.7
- Vemos que existen 30 muertes
- tenemos 210 hombres y 187 mujeres

Realizamos la frecuencia de mis muertes y censuras

Frecuencia de las censuras/fallas de la muestra



```
## >>> suggestions
## piechart(censura_name, hole=0) # traditional pie chart
## piechart(censura_name, values="%") # display %'s on the chart
## piechart(censura_name) # bar chart
## plot(censura_name) # bubble plot
## plot(censura_name, values="count") # lollipop plot
##
```



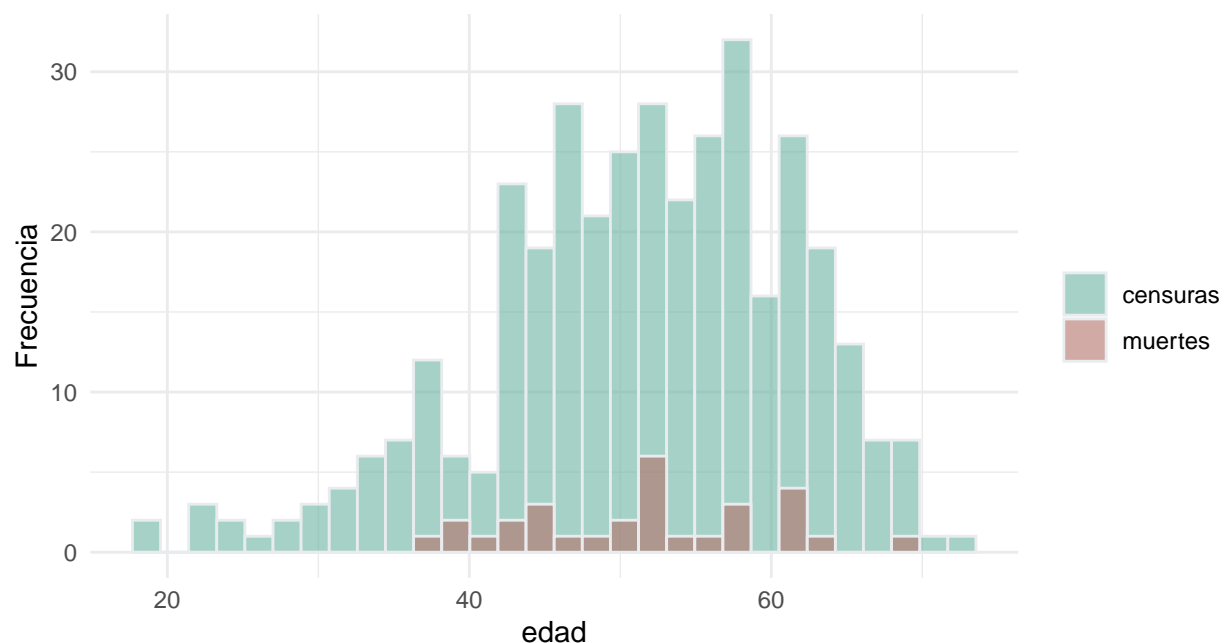
```
## --- censura_name ---
##
##          censuras  muertes    Total
## Frequencies:      367      30     397
## Proportions:    0.924    0.076    1.000
##
## Chi-squared test of null hypothesis of equal probabilities
##   Chisq = 286.068, df = 1, p-value = 0.000
```

Contamos con 367 censuras y 30 muertes

Realizamos la distribución de la edad de los pacientes por censura y muerte :

Distribución de la edad por censura o muerte

Pacientes en una lista de espera de trasplante de hígado de 1990–1999



Kim WR, Therneau TM, Benson JT, Kremers WK, Rosen CB, Gores GJ, Dickson ER. Muertes en la lista de espera de trasplante de hígado: un análisis de riesgos competitivos. Hepatología 2006 Feb; 43(2):345–51.

Observamos que la edad se distribuye con mayor frecuencia entre los 40 y 60 años

Obteniendo la edad promedio de censura y muerte...

```
## [1] "La edad promedio de censura es de: 51.338"
```

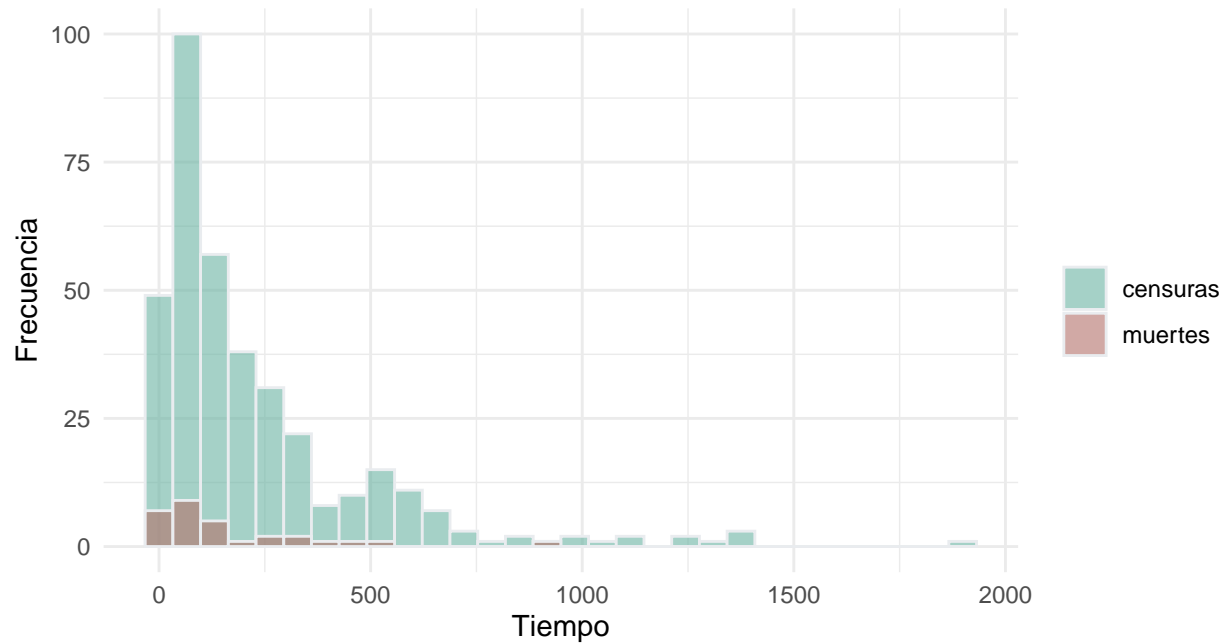
```
## [1] "La edad promedio de muerte es de: 51.7"
```

Como observamos en la gráfica podemos apreciar que el promedio se encuentra en 51

Realizamos la distribución del tiempo por censura o muertes

Distribución del tiempo por censura o muerte

Pacientes en una lista de espera de trasplante de hígado de 1990–1999



Kim WR, Therneau TM, Benson JT, Kremers WK, Rosen CB, Gores GJ, Dickson ER. Muertes en la lista de espera de trasplante de hígado: un análisis de riesgos competitivos. Hepatología 2006 Feb; 43(2):345–51.

Observamos que el tiempo se distribuye con mayor frecuencia entre 0 y 300

Obteniendo el tiempo promedio de censura y muerte...

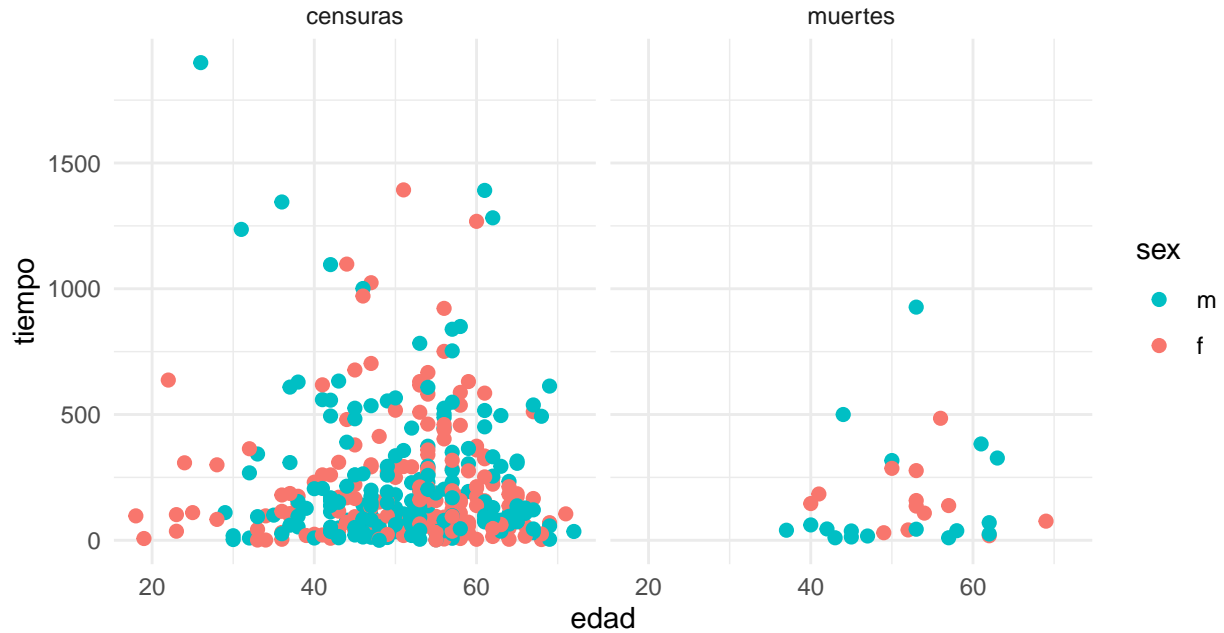
```
## [1] "El tiempo promedio de censura es de: 229.616"
```

```
## [1] "El tiempo promedio de muerte es de: 164.9"
```

Realizamos un gráfico de dispersión de dos módulos que compare el tiempo vs edad de los pacientes por sexo

Edad frente al tiempo de espera de los pacientes por sexo

Pacientes en una lista de espera de trasplante de hígado de 1990–1999

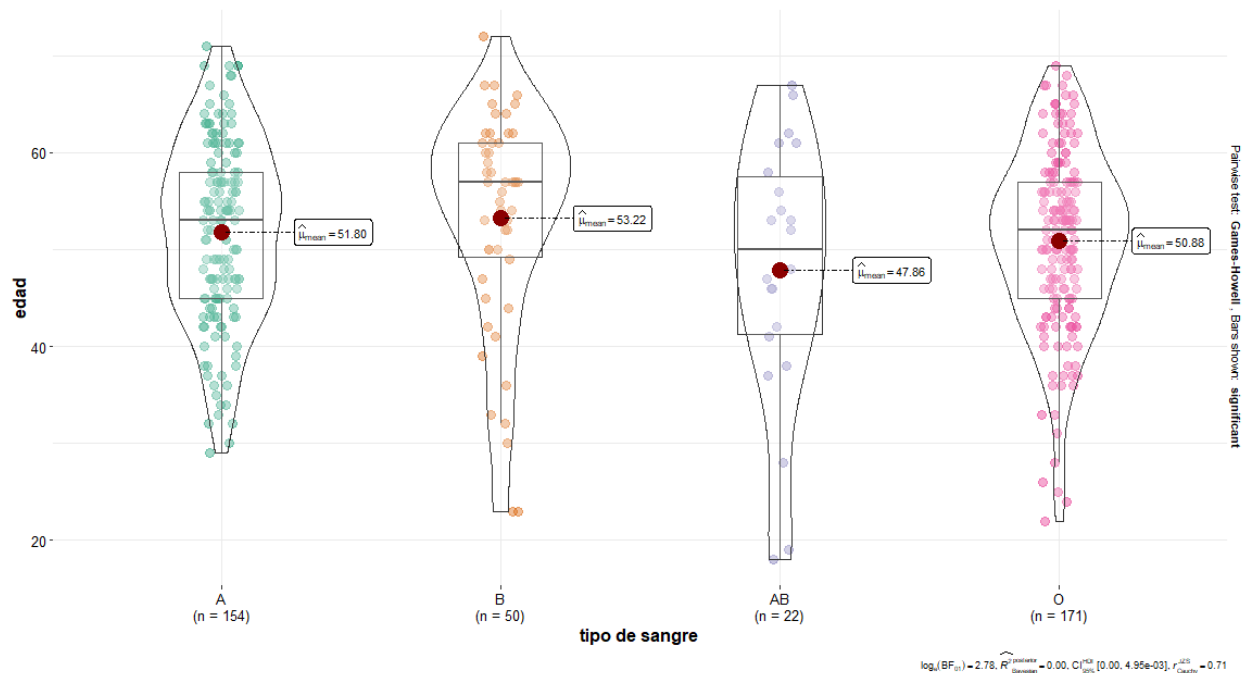


Kim WR, Therneau TM, Benson JT, Kremers WK, Rosen CB, Gores GJ, Dickson ER. Muertes en la lista de espera de trasplante de hígado: un análisis de riesgos competitivos. *Hepatología* 2006 Feb; 43(2):345–51.

vemos que contamos con una menor dispersión en las muertes y no encontramos patrones

Distribución de la edad por tipo de sangre

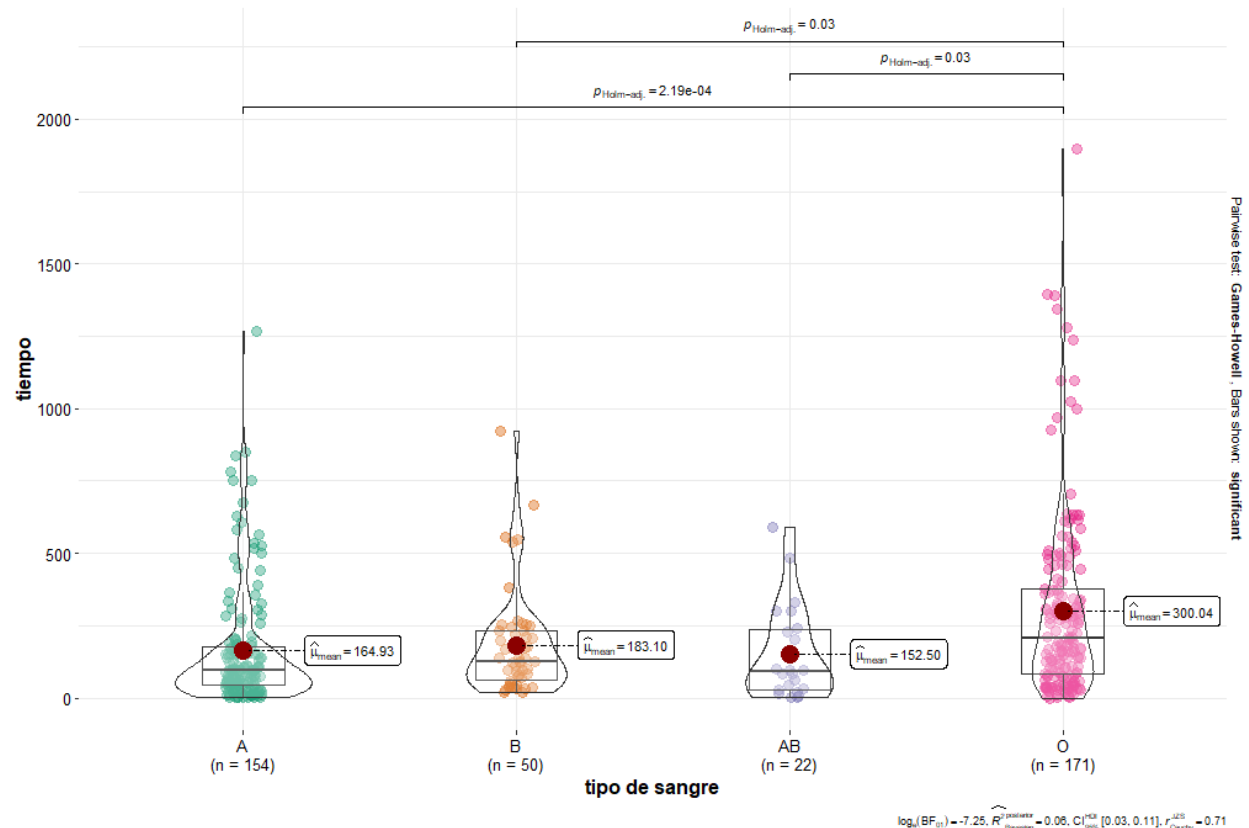
$F_{\text{Welch}}(3, 75.99) = 1.16, p = 0.33, \hat{\omega}_p^2 = 5.91\text{e-}03, \text{CI}_{95\%} [0.00, 1.00], n_{\text{obs}} = 397$



Mediante el boxplot sobre la distribución de la edad por tipo de sangre podemos observar que tenemos un mayor volumen de datos en el tipo de sangre A y O, así como el promedio de edad se encuentra entre 47 y 53 años.

Distribución del tiempo por tipo de sangre

$F_{\text{Welch}}(3, 88.57) = 7.82, p = 1.09\text{e-}04, \hat{\omega}_p^2 = 0.18, \text{CI}_{95\%} [0.06, 1.00], n_{\text{obs}} = 397$



Por otro lado si comparamos el tiempo con el tipo de sangre vemos datos más dispersos a comparación de la edad, se ve que existe una dispersión mayor en el tipo de sangre O.

3.3 ANÁLISIS DE REGRESIÓN

Para el análisis de regresión usaremos la regresión logística siendo esta utilizada para predecir resultados de una variable categórica (una variable que puede adoptar un umero limitado de categorías) en función de las variables independientes o predictoras.

Realizando la regresión logística mediante:

censura ~ age+sex+abo+year

```
##
## Call:
## glm(formula = censura ~ age + sex + abo + year + futime, family = binomial,
##      data = df)
```


Table 1: Interpretación de mis coeficientes

Coeficiente	Interpretación
El coeficiente es igual a 1	La probabilidad de éxito es igual a la de fracaso
El coeficiente es mayor a 1	La probabilidad de éxito es mayor a la de fracaso
El coeficiente es menor a 1	La probabilidad de éxito es menor a la de fracaso

Enfocándonos en la variable “aboB”, siendo esta mi única variable significativa, como es mayor a 1 quiere decir que existe una asociación positiva con la variable, es decir, que existe mayor probabilidad de supervivencia

Probabilidades estimadas por individuo (mostramos solo los primeros items)

```
##          2          4          6          8          10          12
## 0.07429329 0.06515298 0.08757729 0.11455803 0.06281642 0.04935759
```

matriz de confusión

```
##      actual
## pred   0   1
##      0 367  30
```

```
## [1] "nos equivocamos el 7.5566750629723% de las veces"
```

3.4 ANÁLISIS DE SUPERVIVENCIA

3.4.1 MODELO COX

Realizando nuestro modelo COX general en donde se incluye todas las variables:

censura ~ age+sex+abo+year

```
## Call:
## coxph(formula = resp ~ df$age + df$sex + df$abo + df$year)
##
##      n= 397, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## df$age      0.00129   1.00129  0.01858  0.069 0.944654
## df$sexf    -0.21597   0.80576  0.37581 -0.575 0.565511
## df$aboB     0.81510   2.25939  0.50116  1.626 0.103857
## df$aboAB   -0.16746   0.84581  1.06115 -0.158 0.874607
## df$abo0    -0.35974   0.69786  0.45032 -0.799 0.424371
## df$year    -0.27387   0.76043  0.07665 -3.573 0.000353 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
```

```
## df$age      1.0013      0.9987      0.9655      1.0384
## df$sexf     0.8058      1.2411      0.3858      1.6830
## df$aboB     2.2594      0.4426      0.8461      6.0336
## df$aboAB    0.8458      1.1823      0.1057      6.7691
## df$aboO     0.6979      1.4330      0.2887      1.6869
## df$year     0.7604      1.3150      0.6544      0.8837
##
## Concordance= 0.712 (se = 0.047 )
## Likelihood ratio test= 16.3 on 6 df, p=0.01
## Wald test           = 17.5 on 6 df, p=0.008
## Score (logrank) test = 18.51 on 6 df, p=0.005
```

HIPOTESIS:

H0: Los coeficientes de las variables son todos 0 vs ***H1:*** Al menos uno de los coeficientes es diferente de 0
 No rechazo H0, si $p\text{-valor} < \alpha$, rechazamos H0, si al menos uno es mayor a α con un α de 0.05

- *age:* $p\text{-valor} = 0.9446 < \alpha = 0.05?$, No
- *sexf:* $p\text{-valor} = 0.5655 < \alpha = 0.01?$, No
- *aboB:* $p\text{-valor} = 0.1038 < \alpha = 0.01?$, No
- *aboAB:* $p\text{-valor} = 0.8746 < \alpha = 0.01?$, No
- *aboO:* $p\text{-valor} = 0.4243 < \alpha = 0.01?$, No
- *year:* $p\text{-valor} = 0.0003 < \alpha = 0.01?$, Si

Decimos con 95% de confianza que Rechazo H0, al menos uno de los coeficientes es diferente de 0

Podemos decir que al menos una de mis variables influye en la supervivencia del paciente

```
## Call:
## coxph(formula = resp ~ df$age + df$sexf + df$abo + df$year)
##
## n= 397, number of events= 30
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## df$age    0.00129   1.00129  0.01858  0.069 0.944654
## df$sexf   -0.21597   0.80576  0.37581 -0.575 0.565511
## df$aboB    0.81510   2.25939  0.50116  1.626 0.103857
## df$aboAB  -0.16746   0.84581  1.06115 -0.158 0.874607
## df$aboO   -0.35974   0.69786  0.45032 -0.799 0.424371
## df$year   -0.27387   0.76043  0.07665 -3.573 0.000353 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## df$age      1.0013      0.9987      0.9655      1.0384
## df$sexf     0.8058      1.2411      0.3858      1.6830
## df$aboB     2.2594      0.4426      0.8461      6.0336
## df$aboAB    0.8458      1.1823      0.1057      6.7691
## df$aboO     0.6979      1.4330      0.2887      1.6869
## df$year     0.7604      1.3150      0.6544      0.8837
```

```

##
## Concordance= 0.712 (se = 0.047 )
## Likelihood ratio test= 16.3 on 6 df, p=0.01
## Wald test = 17.5 on 6 df, p=0.008
## Score (logrank) test = 18.51 on 6 df, p=0.005

## Call:
## coxph(formula = resp ~ df$sex + df$abo + df$year)
##
## n= 397, number of events= 30
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## df$sexf -0.21618  0.80559  0.37581 -0.575 0.565131
## df$aboB  0.81466  2.25840  0.50097  1.626 0.103913
## df$aboAB -0.17054  0.84321  1.06019 -0.161 0.872206
## df$abo0 -0.36460  0.69447  0.44494 -0.819 0.412532
## df$year -0.27372  0.76055  0.07665 -3.571 0.000356 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## df$sexf  0.8056  1.2413  0.3857  1.6827
## df$aboB  2.2584  0.4428  0.8460  6.0287
## df$aboAB 0.8432  1.1859  0.1056  6.7355
## df$abo0  0.6945  1.4399  0.2904  1.6611
## df$year  0.7605  1.3148  0.6545  0.8838
##
## Concordance= 0.714 (se = 0.046 )
## Likelihood ratio test= 16.3 on 5 df, p=0.006
## Wald test = 17.48 on 5 df, p=0.004
## Score (logrank) test = 18.48 on 5 df, p=0.002

## Call:
## coxph(formula = resp ~ df$abo + df$year)
##
## n= 397, number of events= 30
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## df$aboB  0.78541  2.19330  0.49837  1.576 0.11504
## df$aboAB -0.24708  0.78108  1.05210 -0.235 0.81433
## df$abo0 -0.40293  0.66836  0.43963 -0.917 0.35939
## df$year -0.27140  0.76231  0.07709 -3.521 0.00043 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## df$aboB  2.1933  0.4559  0.82582  5.8252
## df$aboAB 0.7811  1.2803  0.09935  6.1410
## df$abo0  0.6684  1.4962  0.28235  1.5821
## df$year  0.7623  1.3118  0.65541  0.8866
##
## Concordance= 0.707 (se = 0.048 )
## Likelihood ratio test= 15.96 on 4 df, p=0.003
## Wald test = 16.91 on 4 df, p=0.002

```



```
## Score (logrank) test = 17.99 on 4 df, p=0.001

## Call:
## coxph(formula = resp ~ df$year)
##
## n= 397, number of events= 30
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## df$year -0.25905  0.77178  0.07674 -3.376 0.000736 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## df$year    0.7718      1.296    0.664    0.897
##
## Concordance= 0.648 (se = 0.055 )
## Likelihood ratio test= 10.72 on 1 df, p=0.001
## Wald test = 11.4 on 1 df, p=0.0007
## Score (logrank) test = 12 on 1 df, p=0.0005
```

Nos quedamos con el modelo

censura ~ year

HIPOTESIS:

H0: Los coeficientes de las variables son todos 0 vs **H1:** Al menos uno de los coeficientes es diferente de 0
 No rechazo H0, si $p\text{-valor} < \alpha$, rechazamos H0, si al menos uno es mayor a α con un α de 0.05

- $year: \hat{p}\text{-valor} = 0.0007 < \alpha = 0.05?$, Si

Decimos con 95% de confianza que no rechazo H0 Los coeficientes de las variables son todos 0

Podemos decir que la variable **year** influyen en la supervivencia respecto a la presencia de exacerbación del paciente

HIPOTESIS MODELO COX Likelihood ratio test:

H0: Mi modelo COX no es significativo vs **H1:** Mi modelo COX es significativo

Rechazo H0, si $p\text{-valor} < \alpha$, rechazamos H0

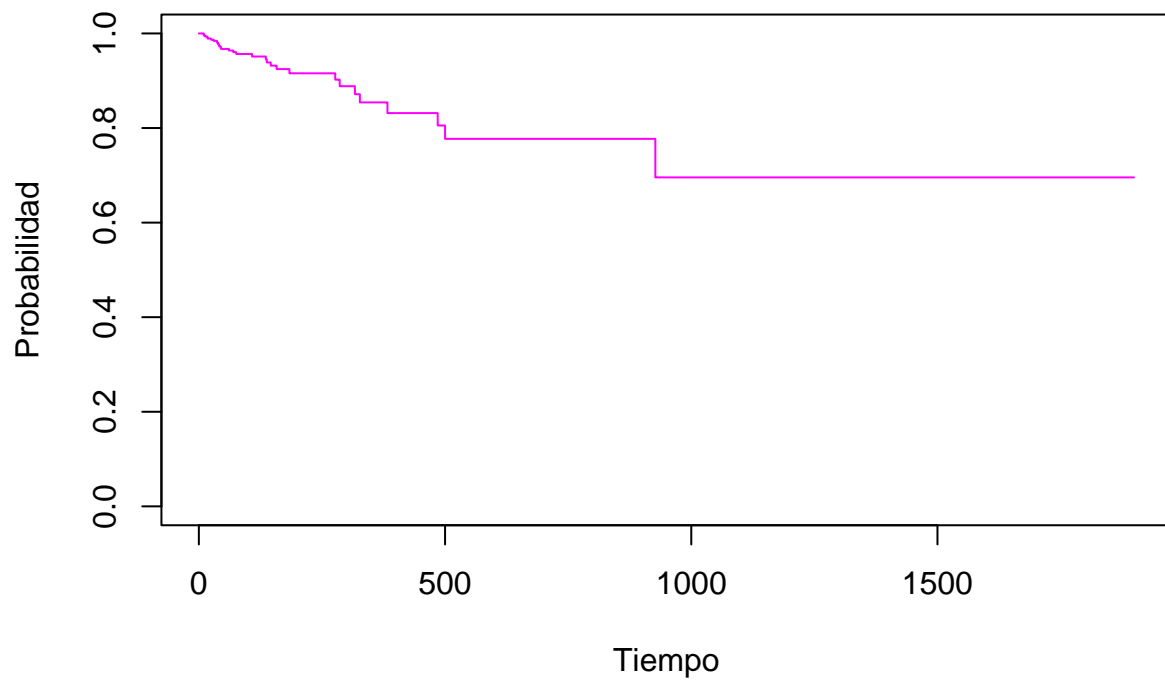
$\hat{p}\text{-valor} = 0.001 < \alpha = 0.05?$, Si

Decimos con 95% de confianza que rechazo H0, mi modelo COX es significativo

```
## Call: survfit(formula = m4)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   10     377      2    0.996 0.00308    0.990    1.000
##   13     372      1    0.994 0.00382    0.986    1.000
##   17     367      1    0.991 0.00447    0.983    1.000
##   18     366      1    0.989 0.00505    0.979    0.999
##   25     353      1    0.987 0.00565    0.976    0.998
##   30     346      1    0.984 0.00625    0.972    0.997
```

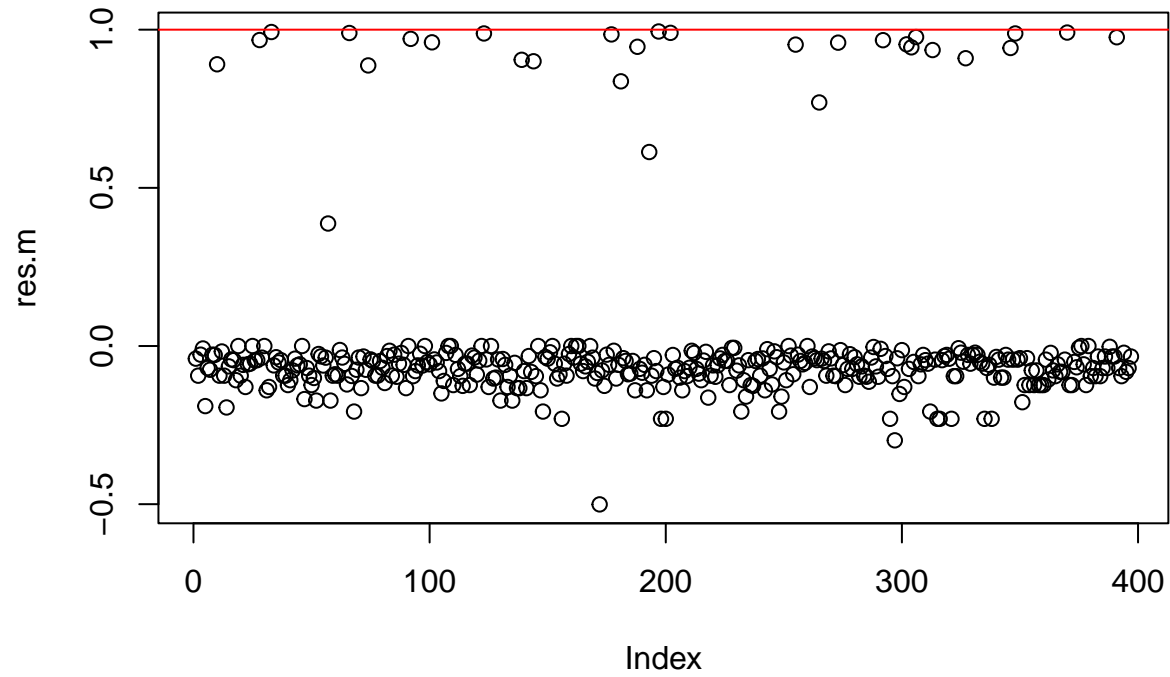
##	37	328	1	0.982	0.00688	0.968	0.995
##	38	325	1	0.979	0.00749	0.964	0.994
##	40	321	1	0.976	0.00806	0.960	0.992
##	41	317	1	0.973	0.00862	0.956	0.990
##	44	314	1	0.970	0.00916	0.952	0.988
##	45	313	1	0.967	0.00967	0.948	0.986
##	61	287	1	0.964	0.01028	0.944	0.984
##	70	278	1	0.960	0.01090	0.939	0.982
##	76	267	1	0.957	0.01155	0.934	0.980
##	108	220	1	0.951	0.01264	0.927	0.976
##	136	201	1	0.945	0.01393	0.918	0.973
##	138	196	1	0.939	0.01517	0.909	0.969
##	146	189	1	0.932	0.01642	0.900	0.965
##	158	176	1	0.925	0.01777	0.890	0.960
##	184	155	1	0.916	0.01963	0.878	0.955
##	277	110	1	0.902	0.02341	0.858	0.949
##	286	105	1	0.888	0.02686	0.837	0.943
##	317	87	1	0.872	0.03135	0.812	0.935
##	327	84	1	0.854	0.03540	0.788	0.926
##	383	69	1	0.832	0.04162	0.754	0.917
##	485	56	1	0.805	0.04861	0.716	0.907
##	500	51	1	0.777	0.05568	0.675	0.894
##	927	13	1	0.696	0.09320	0.535	0.905

Modelo COX



3.4.2 ANÁLISIS DE LOS RESIDUOS DEL MODELO COX

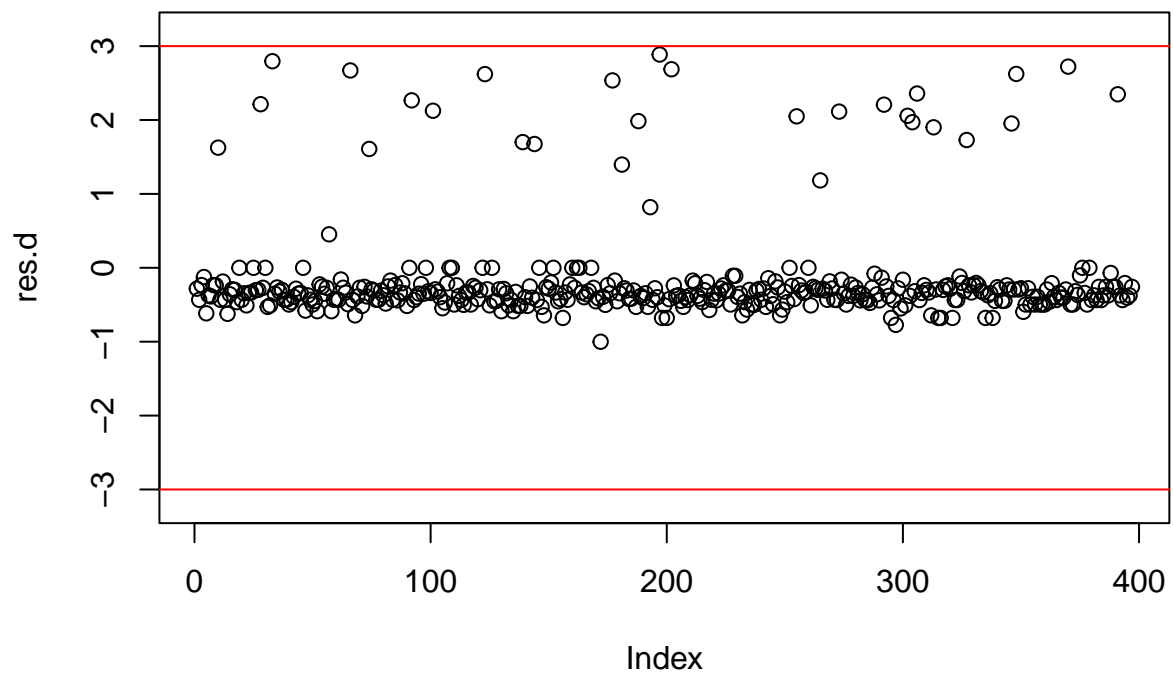
MARTINGALAS



```
## named numeric(0)
```

Los residuos se encuentran dentro de $[-\infty, 1]$

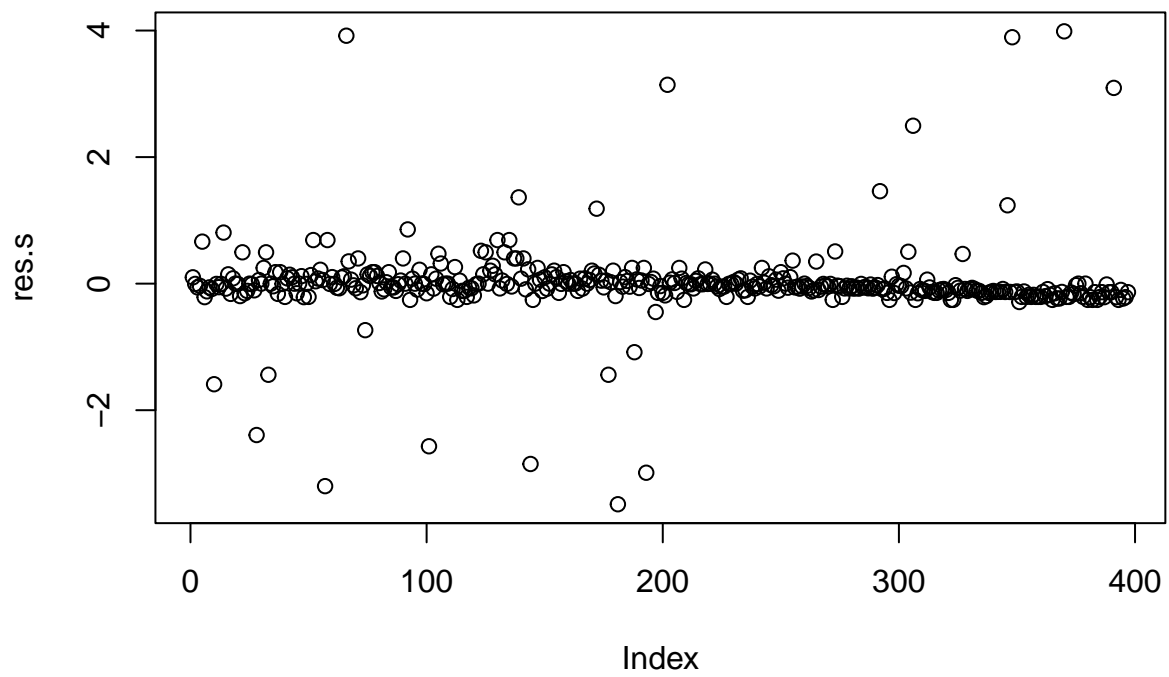
DESVIACION



```
## named numeric(0)
```

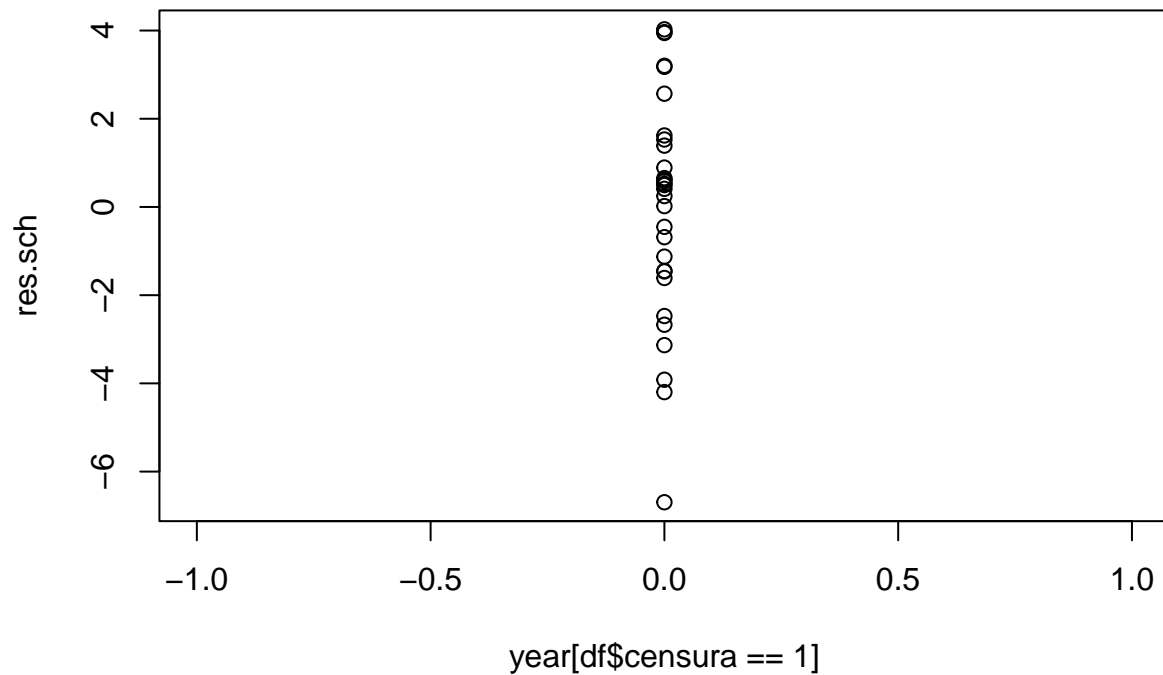
Los residuos se encuentran dentro de $[-3,3]$

SCORE



Los residuos no muestran patrones

SCHOENFELD



Los residuos no muestran patrones ni irregularidades

3.4.2 COMPARACIÓN DEL MODELO COX VS MODELO DE REGRESIÓN

Regresión logística:

censura ~ abo

Modelo COX:

censura ~ year

Comparando los dos modelos, podemos decir que el modelo de regresión logística es más incierto en estos datos comparado con el Modelo COX.

Incluye diferencias en las variables que son tomadas como “mejores”

3.4.3 MODELO KAPLAN-MEIER

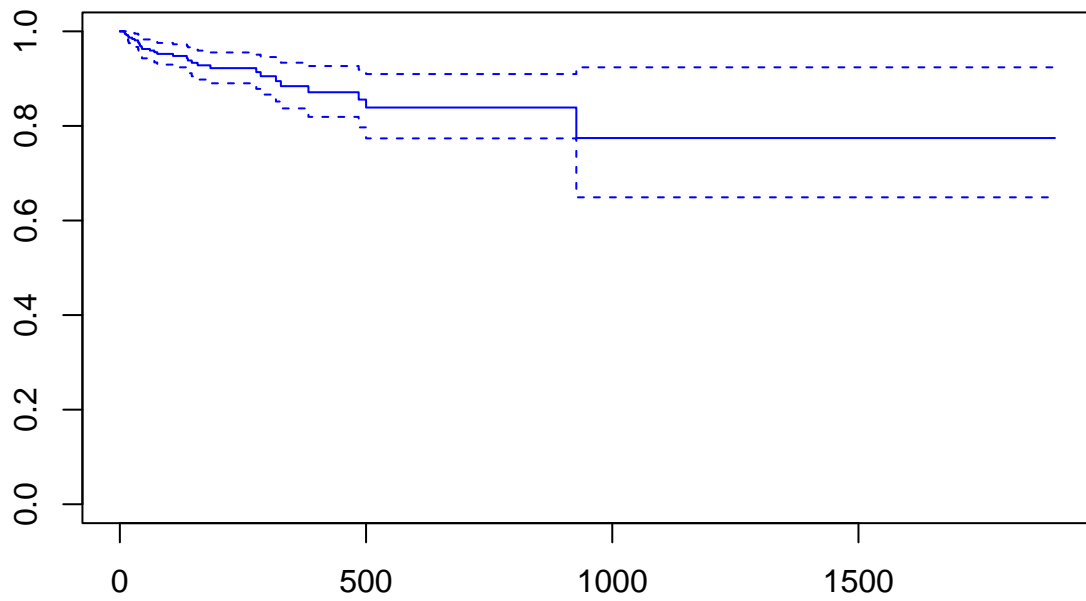
Estimación de la sup modelo KM general:

```
## Call: survfit(formula = resp ~ 1, conf.int = 0.95)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

##	10	377	2	0.995	0.00374	0.987	1.000
##	13	372	1	0.992	0.00459	0.983	1.000
##	17	367	1	0.989	0.00531	0.979	1.000
##	18	366	1	0.987	0.00595	0.975	0.998
##	25	353	1	0.984	0.00655	0.971	0.997
##	30	346	1	0.981	0.00712	0.967	0.995
##	37	328	1	0.978	0.00771	0.963	0.993
##	38	325	1	0.975	0.00825	0.959	0.991
##	40	321	1	0.972	0.00876	0.955	0.989
##	41	317	1	0.969	0.00926	0.951	0.987
##	44	314	1	0.966	0.00973	0.947	0.985
##	45	313	1	0.963	0.01017	0.943	0.983
##	61	287	1	0.959	0.01068	0.939	0.981
##	70	278	1	0.956	0.01118	0.934	0.978
##	76	267	1	0.952	0.01170	0.930	0.976
##	108	220	1	0.948	0.01242	0.924	0.973
##	136	201	1	0.943	0.01323	0.918	0.970
##	138	196	1	0.938	0.01401	0.911	0.966
##	146	189	1	0.933	0.01479	0.905	0.963
##	158	176	1	0.928	0.01562	0.898	0.959
##	184	155	1	0.922	0.01663	0.890	0.955
##	277	110	1	0.914	0.01847	0.878	0.951
##	286	105	1	0.905	0.02024	0.866	0.946
##	317	87	1	0.895	0.02253	0.852	0.940
##	327	84	1	0.884	0.02465	0.837	0.934
##	383	69	1	0.871	0.02742	0.819	0.927
##	485	56	1	0.856	0.03103	0.797	0.919
##	500	51	1	0.839	0.03466	0.774	0.910
##	927	13	1	0.774	0.06977	0.649	0.924

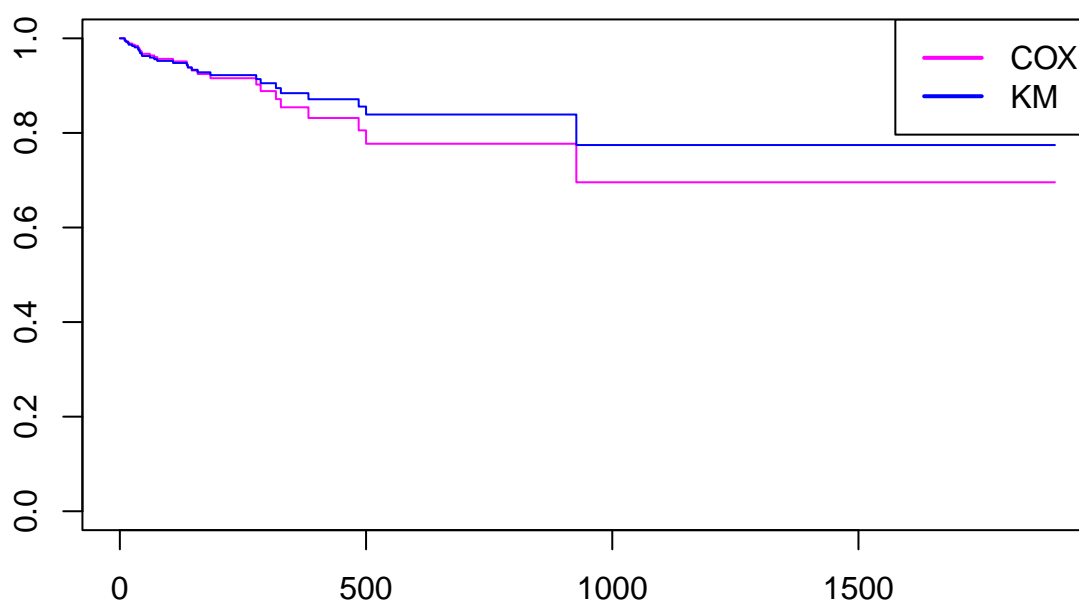
Gráfica de mi modelo general de Kaplan-Meier:

Modelo Kaplan–Meier



Gráfica sobrepuesta de mi estimación COX y KM:

Modelo COX vs Kaplan–Meier(General)



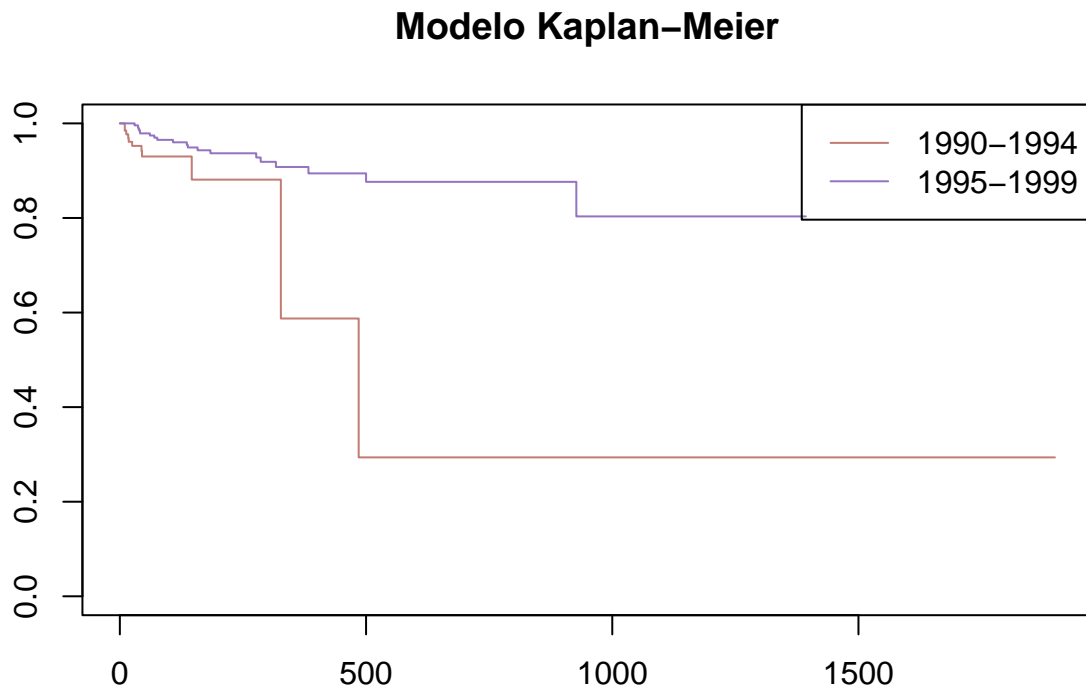
Estimación de la supervivencia modelo KM para la variable “year” donde separamos en dos categorías:

- 1990 - 1994
- 1995 - 1999

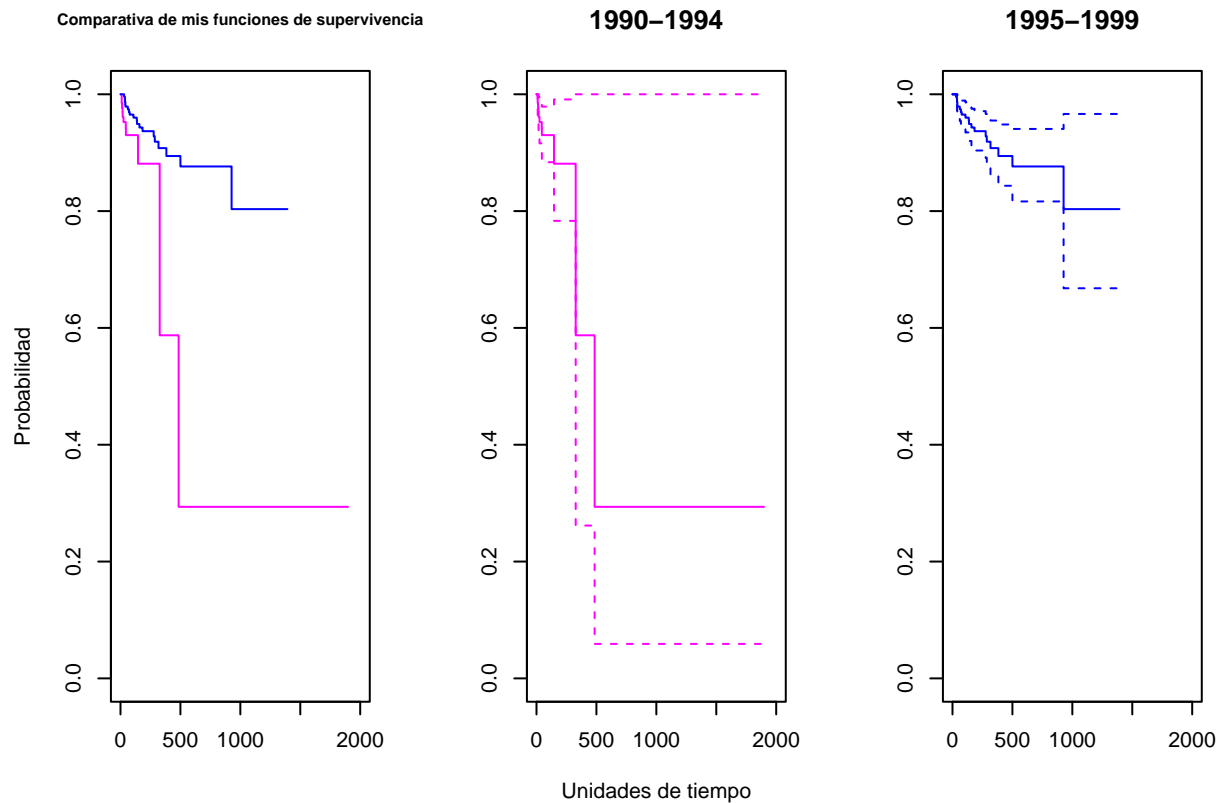
```
## Call: survfit(formula = resp ~ df$year_cat, conf.int = 0.95)
##
##               df$year_cat=1990-1994
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   10   131     2    0.985  0.0107    0.964    1.000
##   13   126     1    0.977  0.0132    0.951    1.000
##   17   123     1    0.969  0.0153    0.939    0.999
##   18   122     1    0.961  0.0171    0.928    0.995
##   25   112     1    0.952  0.0190    0.916    0.990
##   44    85     1    0.941  0.0218    0.899    0.985
##   45    84     1    0.930  0.0243    0.884    0.979
##  146    19     1    0.881  0.0529    0.783    0.991
##  327     3     1    0.587  0.2424    0.262    1.000
##  485     2     1    0.294  0.2405    0.059    1.000
##
##               df$year_cat=1995-1999
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   30   241     1    0.996  0.00414    0.988    1.000
##   37   235     1    0.992  0.00591    0.980    1.000
##   38   234     1    0.987  0.00724    0.973    1.000
##   40   233     1    0.983  0.00836    0.967    1.000
```

##	41	231	1	0.979	0.00934	0.961	0.997
##	61	218	1	0.974	0.01032	0.954	0.995
##	70	212	1	0.970	0.01125	0.948	0.992
##	76	209	1	0.965	0.01212	0.942	0.989
##	108	187	1	0.960	0.01311	0.935	0.986
##	136	177	1	0.955	0.01411	0.927	0.983
##	138	174	1	0.949	0.01506	0.920	0.979
##	158	159	1	0.943	0.01610	0.912	0.975
##	184	149	1	0.937	0.01719	0.904	0.971
##	277	106	1	0.928	0.01917	0.891	0.966
##	286	101	1	0.919	0.02107	0.878	0.961
##	317	84	1	0.908	0.02348	0.863	0.955
##	383	67	1	0.894	0.02676	0.843	0.948
##	500	50	1	0.876	0.03164	0.817	0.941
##	927	12	1	0.803	0.07570	0.668	0.966

Gráfica de mi modelo general de Kaplan-Meier:



Gráfica 3-1 de la variable year (1990-1994 , 1995-1999:



Prueba Analítica:

Comparacion de funciones de supervivencia

HIPOTESIS:

H0: Las funciones de supervivencia son iguales H1: Las funciones de supervivencia al menos una difiere

```
## Call:
## survdiff(formula = resp ~ df$year_cat, rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## df$year_cat=1990-1994 146      11    5.59    5.24    7.11
## df$year_cat=1995-1999 251      19   24.41    1.20    7.11
##
## Chisq= 7.1  on 1 degrees of freedom, p= 0.008
```

Rechazo H0 si pvalor: $0.008 < \alpha: 0.05$, no rechazo

Con 95% de confianza no rechazo H0, funciones de supervivencia iguales

apoyados en el gráfico, los del año 1995-1999 tienen mayor probabilidad de supervivencia

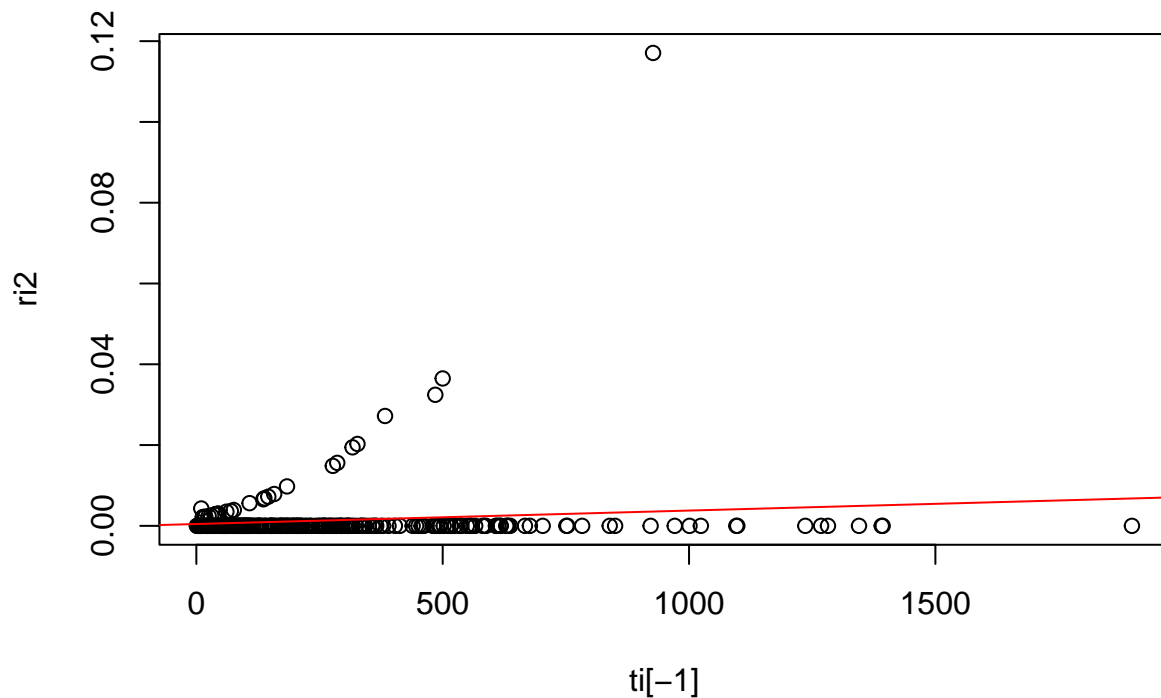
3.4.4 MODELOS PARAMÉTRICOS

EXPONENCIAL

Revisión de la “pista” gráfica para asociar a distribución exponencial, riesgo vs tiempo, buscamos linea recta horizontal:

```
##
## Call:
## lm(formula = ri2 ~ ti[-1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.006779 -0.001646 -0.001026 -0.000645  0.113538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.000451333 0.000698317   0.646   0.5186
## ti[-1]      0.000003332 0.000001706   1.953   0.0519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008206 on 264 degrees of freedom
## Multiple R-squared:  0.01424,    Adjusted R-squared:  0.01051
## F-statistic: 3.815 on 1 and 264 DF,  p-value: 0.05185
```

Modelo Exponencial



vemos que no se cumple la linea horizontal

HIPOTESIS:

H0: El modelo es una linea horizontal vs H1: El modelo no es una linea horizontal

Rechazo H0 si $\text{coe}[\text{ti}[-1]]$ es diferente de 0, si rechazo

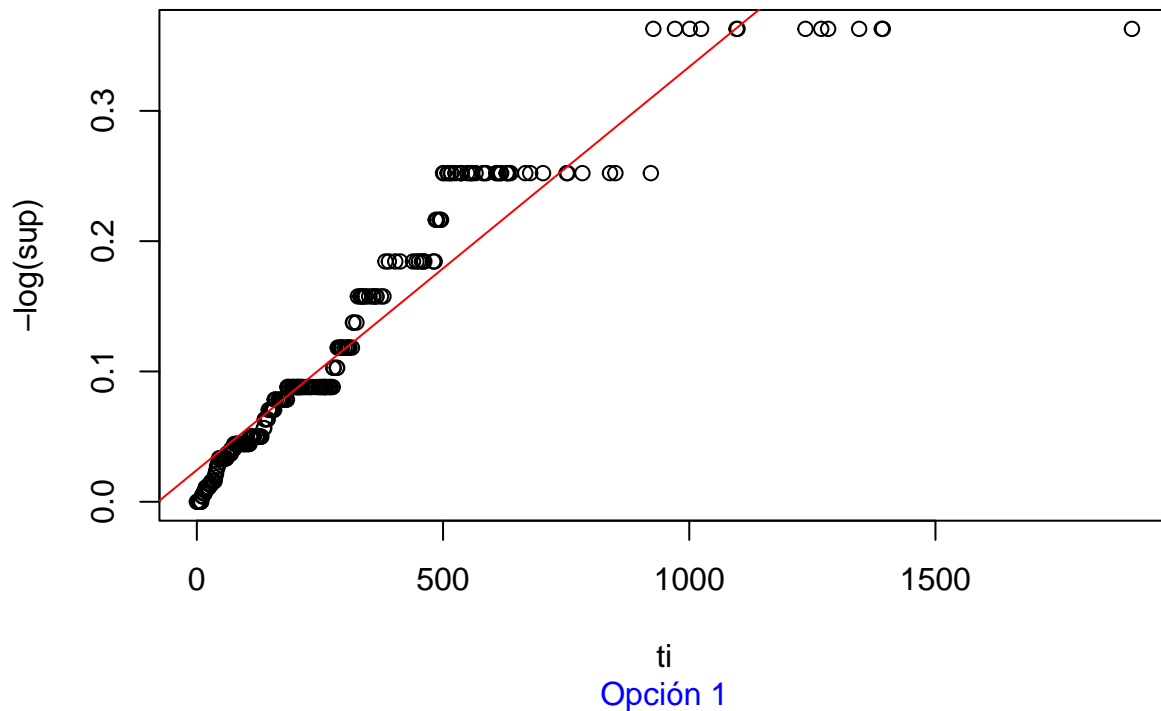
La regresión es significativa, no encontramos recta horizontal

WEIBULL

Opción 1. buscamos recta al origen en $-\log(\text{sup})$ vs tiempo:

```
##
## Call:
## lm(formula = -log(sup) ~ ti)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.249054 -0.012621 -0.005312  0.008528  0.073114
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) 0.024374161 0.002534542   9.617 <0.0000000000000002 ***
## ti          0.000309424 0.000006203  49.881 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02989 on 265 degrees of freedom
## Multiple R-squared:  0.9037, Adjusted R-squared:  0.9034
## F-statistic: 2488 on 1 and 265 DF, p-value: < 0.00000000000000022
```

Modelo Weibull sin regresión al origen



HIPOTESIS:

H_0 : El modelo no es significativo (variables iguales $B_1=B_2=B_3=\dots=B_K$)

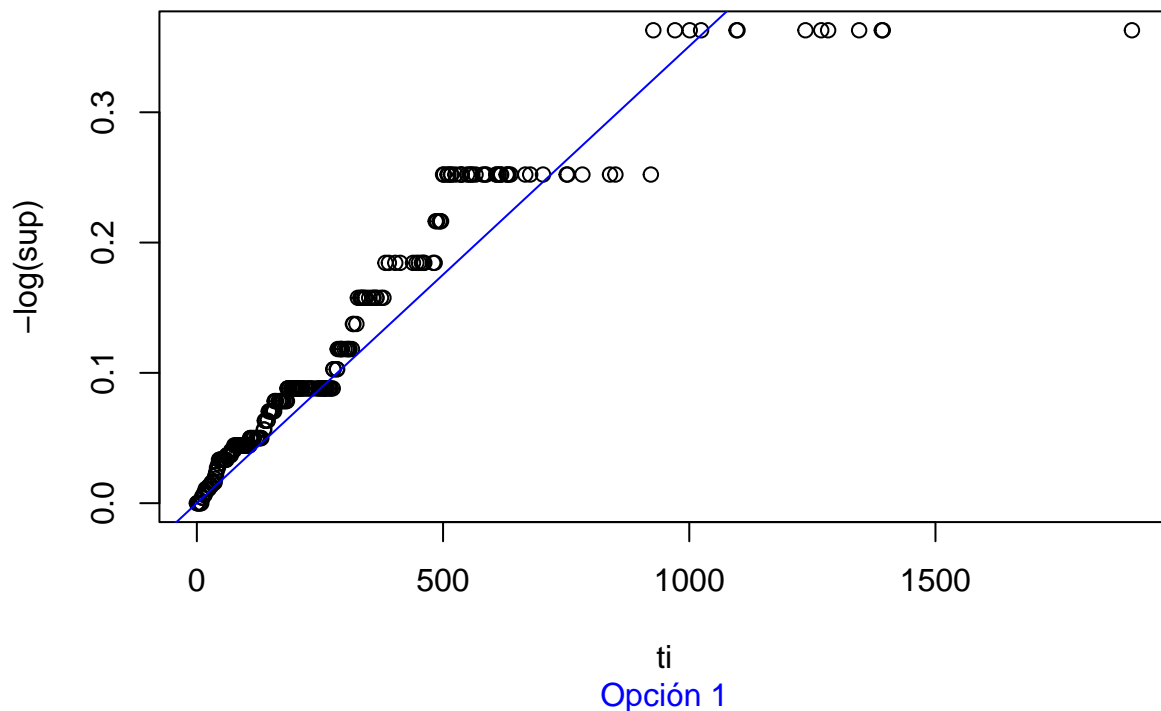
H_1 : El modelo es significativo (al menos un coeficiente diferente de 0)

Rechazo H_0 si p valor: $0.00000000000000022 < \alpha: 0.05$, si rechazo

La regresión es significativa, con un ajuste del 90.34%

```
##
## Call:
## lm(formula = -log(sup) ~ ti + 0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.303095  0.005185  0.014205  0.021305  0.076841
##
## Coefficients:
##      Estimate Std. Error t value      Pr(>|t|)
## ti 0.00035072 0.00000519   67.58 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03465 on 266 degrees of freedom
## Multiple R-squared:  0.945, Adjusted R-squared:  0.9447
## F-statistic: 4566 on 1 and 266 DF, p-value: < 0.00000000000000022
```

Modelo Weibull con regresión al origen



HIPOTESIS:

H_0 : El modelo no es significativo (variables iguales $B_1=B_2=B_3=\dots=B_K$)

H_1 : El modelo es significativo (al menos un coeficiente diferente de 0)

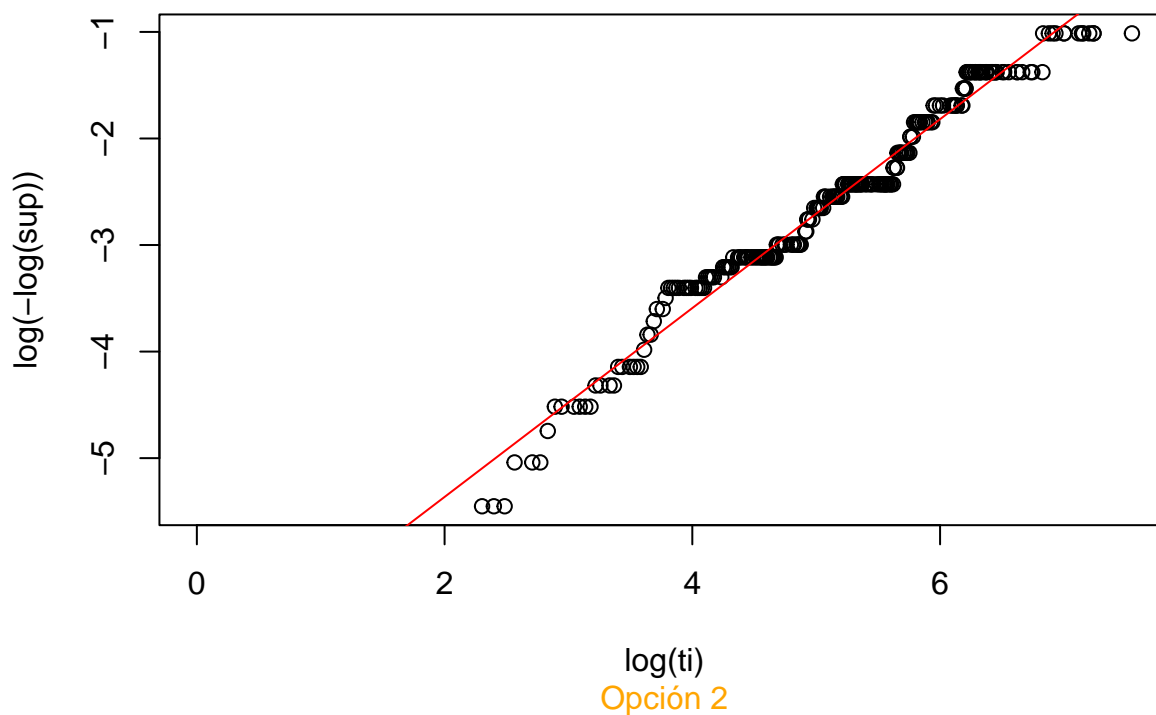
Rechazo H_0 si p valor: $0.00000000000000022 < \alpha: 0.05$, si rechazo

La regresión es significativa, con un ajuste del 94.47%

Opción 2. buscamos recta en $\log(-\log(\text{sup}))$ vs $\log(\text{tiempo})$:

```
##
## Call:
## lm(formula = log(-log(sup))[10:length(sup)] ~ log(ti)[10:length(sup)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56726 -0.08934  0.01131  0.10530  0.36083
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    -7.136810   0.045349  -157.4 <0.0000000000000002 ***
## log(ti)[10:length(sup)]  0.886265   0.008569   103.4 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1485 on 256 degrees of freedom
## Multiple R-squared:  0.9766, Adjusted R-squared:  0.9765
## F-statistic: 1.07e+04 on 1 and 256 DF, p-value: < 0.00000000000000022
```

Modelo Weibull



HIPOTESIS:

H_0 : El modelo no es significativo (variables iguales $B_1=B_2=B_3=\dots=B_K$)

H1: El modelo es significativo (al menos un coeficiente diferente de 0)

Rechazo H_0 si p valor: 0.00000000000000022 < alfa: 0.05, si rechazo

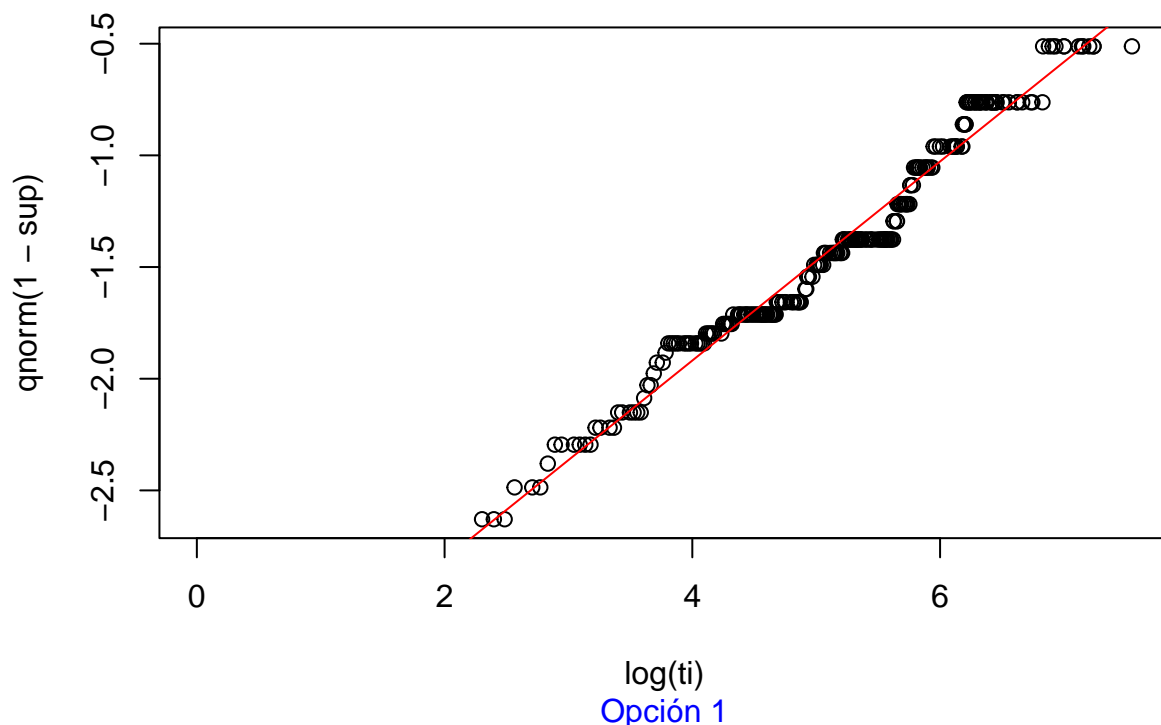
La regresión es significativa, con un ajuste del 97.65%, el modelo Weibull es una sugerencia.

LOG NORMAL

Opción 1. usamos la inversa de normal estándar:

```
##
## Call:
## lm(formula = qnorm(1 - sup)[10:length(sup)] ~ log(ti)[10:length(sup)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.180103 -0.051216 -0.002401  0.057561  0.169455
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    -3.70101     0.02419  -153.03 <0.0000000000000002 ***
## log(ti)[10:length(sup)]  0.44559     0.00457   97.51 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0792 on 256 degrees of freedom
## Multiple R-squared:  0.9738, Adjusted R-squared:  0.9737
## F-statistic: 9508 on 1 and 256 DF, p-value: < 0.00000000000000022
```

Modelo Log Normal



HIPOTESIS:

H0: El modelo no es significativo (variables iguales $B_1=B_2=B_3=\dots=B_K$)

H1: El modelo es significativo (al menos un coeficiente diferente de 0)

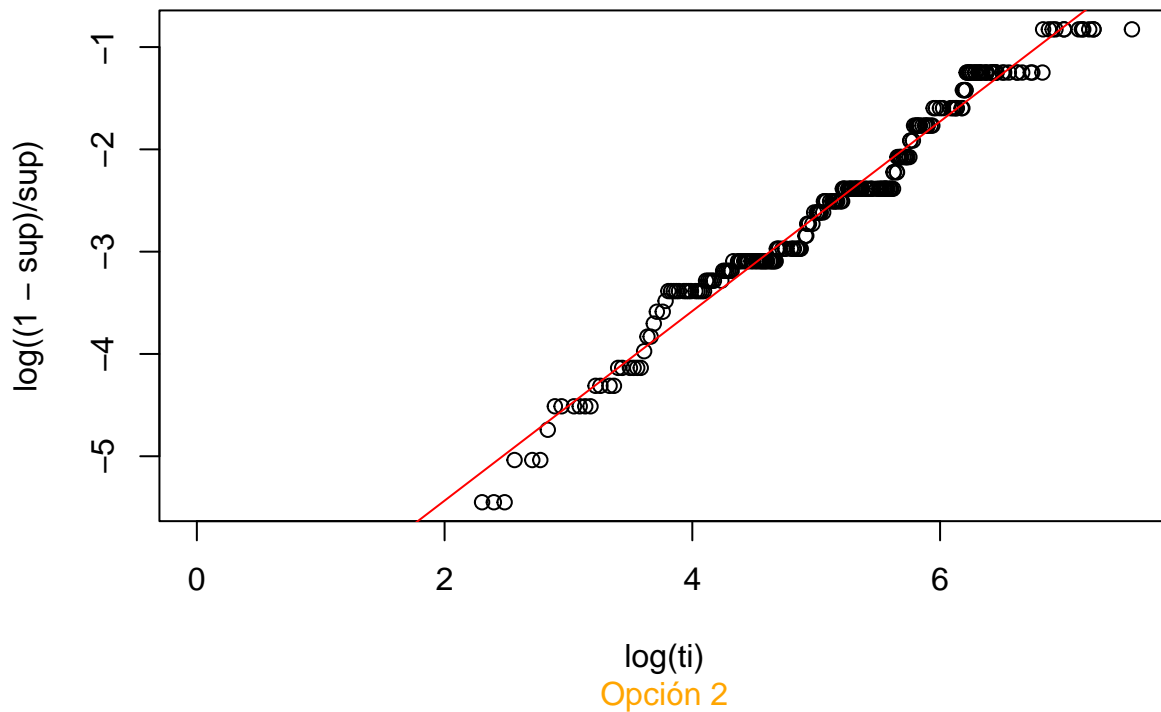
Rechazo H0 si p valor: 0.00000000000000022 < alfa: 0.05, si rechazo

La regresión es significativa, con un ajuste del 97.37%

Opción 2.:

```
##
## Call:
## lm(formula = log((1 - sup)/sup)[10:length(sup)] ~ log(ti)[10:length(sup)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5348 -0.1002  0.0048  0.1085  0.3742
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)    -7.287315    0.046188  -157.8 <0.0000000000000002 ***
## log(ti)[10:length(sup)]  0.926669    0.008727   106.2 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1513 on 256 degrees of freedom
## Multiple R-squared:  0.9778, Adjusted R-squared:  0.9777
## F-statistic: 1.127e+04 on 1 and 256 DF,  p-value: < 0.00000000000000022
```

Modelo Log Normal



HIPOTESIS:

H0: El modelo no es significativo (variables iguales $B_1=B_2=B_3=\dots=B_K$)

H1: El modelo es significativo (al menos un coeficiente diferente de 0)

Rechazo H0 si p valor: $0.00000000000000022 < \alpha: 0.05$, si rechazo

La regresión es significativa, con un ajuste del 97.77%, el modelo LOG NORMAL es una sugerencia

3.4.4 COMPARACIÓN DE LOS MODELOS PARAMETRICOS

Revisamos los modelos sugeridos

WEIBULL

```
## Call:
## survreg(formula = resp ~ df$year[df$futime > 0], dist = "weibull")
##
## Coefficients:
##           (Intercept) df$year[df$futime > 0]
##          -576.2697121             0.2928153
##
## Scale= 1.084604
##
## Loglik(model)= -262   Loglik(intercept only)= -267.9
##  Chisq= 11.84 on 1 degrees of freedom, p= 0.00058
## n= 396
```

HIPOTESIS:

H0: El modelo no es significativo ($B1=0$)

H1: El modelo es significativo ($B1$ diferente de 0)

Rechazo H0 si p valor: $0.00058 < \alpha: 0.05$, si rechazo

El modelo es significativo

LOG NORMAL

```
## Call:
## survreg(formula = resp ~ df$year[df$futime > 0], dist = "lognormal")
##
## Coefficients:
##             (Intercept) df$year[df$futime > 0]
##             -552.8345800             0.2812299
##
## Scale= 2.200273
##
## Loglik(model)= -261.4   Loglik(intercept only)= -266.6
##   Chisq= 10.38 on 1 degrees of freedom, p= 0.00127
## n= 396
```

HIPOTESIS:

H0: El modelo no es significativo ($B1=0$)

H1: El modelo es significativo ($B1$ diferente de 0)

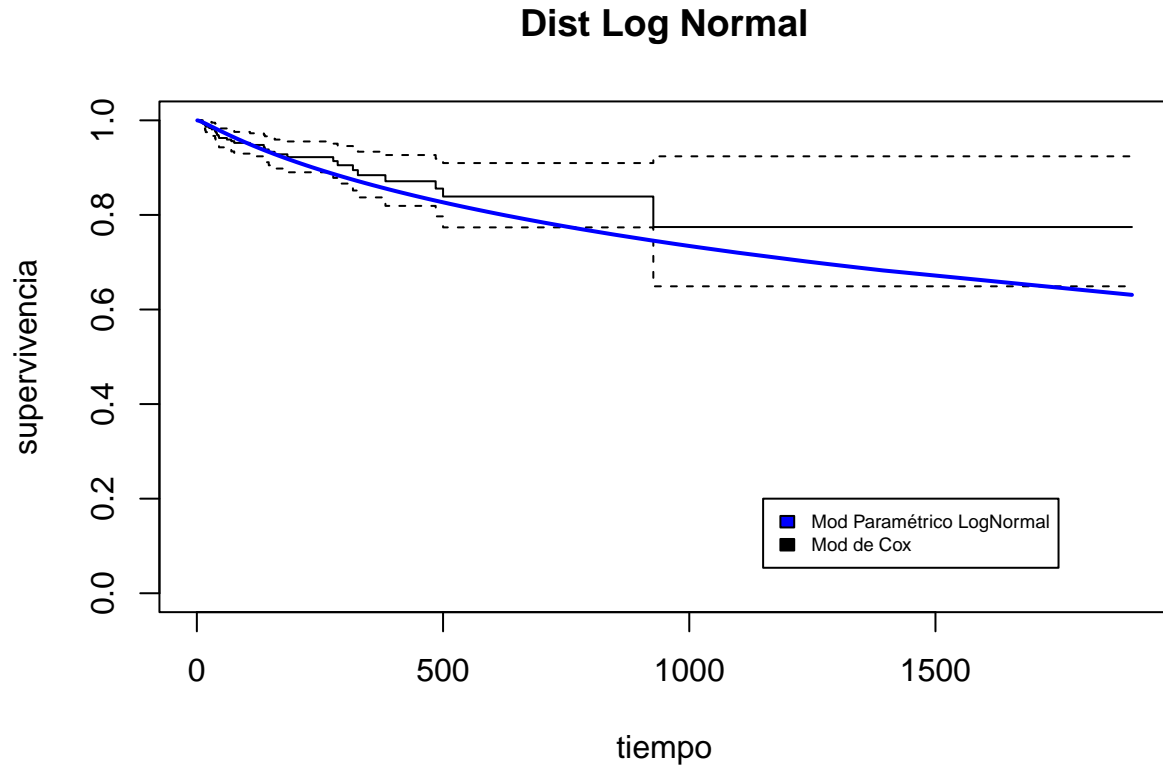
Rechazo H0 si p valor: $p= 0.00127 < \alpha: 0.05$, si rechazo

El modelo es significativo

Realizamos la elección del modelo mediante el parámetro AIC

##	lognormal	weibull	llogis	gompertz
## AIC	528.8297	529.9751	530.0554	540.6841
## BIC	540.7739	541.9194	541.9997	552.6283
## LogLik	-261.4148	-261.9876	-262.0277	-267.3420

Entonces decimos que nuestro mejor modelo paramétrico es el Lognormal



Función de supervivencia:

$$S(t) = 1 - \Phi \frac{\ln(t) - \mu}{\sigma} = 1 - 0.281 \frac{\ln(t) + 552.835}{2.2}$$

Función de densidad:

$$f(t) = \frac{(\Phi \frac{\ln(t) - \mu}{\sigma})}{t} = \frac{(0.281 \frac{\ln(t) - 552.835}{2.2})}{t}$$

Función de riesgo:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{(0.281 \frac{\ln(t) - 552.835}{2.2})}{t}}{1 - 0.281 \frac{\ln(t) + 552.835}{2.2}}$$

4. CONCLUSIÓN

El modelo diseñado sirvió para analizar los datos de la experiencia de la región en particular, y como se sabe durante un período de tiempo en el que el trasplante de hígado se reconoció mucho más ampliamente como una modalidad de tratamiento viable. El número de trasplantes de hígado aumenta durante el período, pero el número de sujetos agregados a la lista de espera de trasplante de hígado creció mucho más rápido. Para ello se contruyó este modelo, para poder seleccionar a los clientes dada las variables de los pacientes, así poder seleccionar mejor y evitar fallos/muertes.

5. BIBLIOGRAFIAS

- *Martinez, J. (s. f.). Análisis de Supervivencia en R. http://rstudio-pubs-static.s3.amazonaws.com/316989_83cbe556125645b698c9ff6cf88c4c1a.html*
- *Kim WR, Therneau TM, Benson JT, Kremers WK, Rosen CB, Gores GJ, Dickson ER. Muertes en la lista de espera de trasplante de hígado: un análisis de riesgos competitivos. Hepatología 2006 Feb; 43(2):345-51.*