



**FACULTAD DE  
CIENCIAS FÍSICO  
MATEMÁTICAS**



**INTEGRANTES:**

**ANA KAREN GÓMEZ ALONZO**

**ALMA CECILIA VILLARREAL  
DUARTE**

**SAHORI VERÓNICA RAMÍREZ  
MEDELLÍN**

**ANGEL EMMANUEL HERNÁNDEZ  
GALVÁN**

**GRUPO:001**

*Estadística Aplicada*

# INDICE

<b><u>Introducción</u></b>	<b><u>3</u></b>
<b><u>Contexto</u></b>	<b><u>3</u></b>
<b><u>Descripción de datos</u></b>	<b><u>3</u></b>
<b><u>Evento de interés</u></b>	<b><u>4</u></b>
<b><u>Descripción de variables</u></b>	<b><u>4</u></b>
<b><u>Objetivos</u></b>	<b><u>7</u></b>
<b><u>Análisis</u></b>	<b><u>7</u></b>
<b><u>Análisis descriptivo de los datos</u></b>	<b><u>7</u></b>
<b><u>Análisis de regresión para selección de variables</u></b>	<b><u>12</u></b>
<b><u>Análisis de supervivencia de forma</u></b>	<b><u>14</u></b>
<b><u>Modelo de Cox</u></b>	<b><u>14</u></b>
<b><u>Método Kaplan-Meier</u></b>	<b><u>20</u></b>
<b><u>Modelos paramétricos</u></b>	<b><u>26</u></b>
<b><u>Comparación de modelos</u></b>	<b><u>30</u></b>
<b><u>Conclusiones</u></b>	<b><u>31</u></b>
<b><u>Bibliografía</u></b>	<b><u>32</u></b>

# Introducción

## Contexto

### Definición de cirrosis

Los conductos biliares son tubos por los cuales la bilis pasa del hígado al intestino delgado. La bilis es una sustancia que ayuda con la digestión. Todos los conductos biliares juntos se denominan tracto biliar.

Cuando los conductos biliares se hinchan o inflaman, esto bloquea el flujo de bilis. Estos cambios pueden conducir a la cicatrización del hígado llamada cirrosis. Esto se llama cirrosis biliar.

*Los valores normales de la bilirrubina son menores a 1 mg/dL.*

### Ursodiol (UDCA)

El Ursodiol es un ácido biliar que disminuye la cantidad de colesterol producido por el hígado y absorbido por los intestinos. Este ayuda a degradar el colesterol que ha formado cálculos en la vesícula biliar. También aumenta el flujo de bilis en pacientes con cirrosis biliar primaria. Especialmente se usa para el tratamiento de la cirrosis biliar primaria, pero de igual manera puede usarse para fines no mencionados en la guía del medicamento.

### Puntuación de riesgo de mayo

Es una herramienta estadística que ayuda a en las predicciones de supervivencia de los pacientes con cirrosis biliar primaria (CBP ó PBC) que no se han realizado alguna operación. La fórmula para calcular el riesgo es la siguiente:

$$\text{Riesgo} = 0.871 \log_e (\text{bilirrubina en mg/dL}) + (-2.53) \log_e (\text{albumina en g/dL}) + 0.039 \text{ edad en años} + 2.38 \log_e (\text{protrombina tiempo en segundos}) + 0.859$$
  
(puntuación de edema de 0, 0.5, o 1).

### Descripción de datos

La base de datos “udca” muestra los resultados obtenidos en las pruebas del tratamiento con UDCA de pacientes con cirrosis biliar primaria. Teniendo un total de 170 observaciones y 15 variables.

Por el otro lado, en la base de datos “udca1” sobre el mismo estudio, se tiene el mismo número de observaciones, pero con 7 variables que simplifican la información encontrada en el “udca”.

## Evento de interés

En las últimas 8 columnas de la base de datos “udca”, se encuentran diversas fechas en las que el paciente sufrió alguna de las siguientes situaciones:

- Muerte.
- Trasplante de hígado.
- Progresión histológica.
- Aparición de varices esofágicas.
- Aparición de ascitis.
- Aparición de encefalopatía.
- Duplicación de bilirrubina inicial.
- Agravamiento de los síntomas.

Todas estas situaciones indican un problema o deterioro del funcionamiento del hígado. El conjunto de datos contiene una cantidad muy considerable de valores NA en estas columnas, esto se debe a que se tomaron las fechas de la aparición de cualquiera de las situaciones y cuando nunca sucedieron se dejó un NA como valor.

Lo anterior se puede observar en la tabla Udca de descripción de variables.

## Descripción de variables

**Tabla Udca**

Variable	Descripción	Escala de medición	Tipo de variable	NA's
Id	identificador	Numérica, tipo entero (de 1 a 170)	Discreta	0
Trt	Tipo de tratamiento	Numérica. 0=placebo, 1=UDCA.	Discreta	0
entry.dt	Fecha en que el paciente entro a estudio.	Fecha.	Continua	0
last.dt	Ultima fecha en la que el paciente estuvo en el estudio.	Fecha.	Continua	0
Stage	Etapas de enfermedad.	Numérico, 0 y 1.	Discreta	0

Bili	Valor de bilirrubina al entrar	Numérico, decimal de 1 cifra. (mg/dL).	Continua	0
Riskscore	La puntuación de riesgo de Mayo	Numérico, decimal de 1 cifra.	Continua	1
death.dt	Fecha de muerte	Fecha.	Continua	154
tx.dt	Fecha en la que se le hizo el trasplante al paciente.	Fecha.	Continua	158
hprogress.dt	Fecha de progresión histológica.	Fecha.	Continua	150
varices.dt	Aparición de varices esofágicas.	Fecha.	Continua	145
ascites.dt	Aparición de ascitis.	Fecha.	Continua	164
enceph.dt	Aparición de encefalopatía	Fecha.	Continua	166
double.dt	Duplicación de bilirrubina inicial.	Fecha.	Continua	153
worsen.dt	Agravamiento de síntomas en 2 etapas.	Fecha.	Continua	154

Este data set contiene a su vez 2 data sets: udca1 y udca2, el primero de ellos contiene exactamente la misma información y sujetos de estudio, sin embargo, sustituye las últimas 8 columnas por una sola llamada “status”, esta variable es booleana y sus valores significan lo siguiente:

1: Si el paciente sufrió al menos uno de los eventos que indican deterioro del funcionamiento del hígado.

0: si no ha padecido ninguno.

A la par hace un cambio para las variables entry.dt y last.dt y las sustituye por una variable numérica llamada “futime”, que indica el tiempo transcurrido desde que el paciente entro en el estudio hasta que se presentó el primer evento que indica deterioro en el funcionamiento del hígado, el cual es nuestro evento de interés.

La descripción de las variables de la data set “udca1” se encuentra en la tabla Udca1. Este data set será el usado para el análisis.

**Tabla Udca1**

Variable	Descripción	Escala de medición	Tipo de variable	NA's
Id	Identificador	Numérica, tipo entero (de 1 a 170)	Discreta	0
Trt	Tipo de tratamiento	Numérica. 0=placebo, 1=UDCA.	Discreta	0
Stage	Etapas de enfermedad.	Numérico, 0 y 1.	Discreta	0
Bili	Valor de bilirrubina al entrar	Numérico, decimal de 1 cifra. (mg/dL).	Continua	0
riskscore	La puntuación de riesgo de Mayo	Numérico, decimal de 1 cifra.	Continua	1
futime	Tiempo en días en el que ocurre el primer suceso que presenta un deterioro al hígado.	Numérico, tipo entero.	Continua	0
Status	Determina si el hígado ha sufrido una falla, ya sea la muerte, trasplante de hígado o alguna de las enfermedades que pudiesen deteriorar su estado.	Numérico, 0= No presenta falla 1=Presenta una falla	Discreta	0

Se realizó una limpieza de datos para eliminar la fila que contiene el NA encontrado, dejando así 169 observaciones y 7 columnas.

## Objetivos

- Determinar si hay diferencia significativa en el mantenimiento de las funciones del hígado entre los individuos que se les dio UDCA como tratamiento y los que tomaron el placebo.
- Determinar que variables son significativas para el mantenimiento de las funciones del hígado.
- Determinar si los puntajes más altos registrados para el riesgo de mayo en los pacientes poseen un mayor o menor índice de supervivencia.

## Análisis

### Análisis descriptivo de los datos

#### Número de eventos de interés

```
estatus
Sin falla Con falla
      97      72
```

De los 169 datos a trabajar, 72 son nuestro evento de interés.

#### Promedio de tiempo de eventos

Tiempo promedio del evento de interés: 742.7778 días.

Tiempo promedio de los censurados: 1250.133 días.

#### Tablas de contingencia

- Por tratamiento

0 = Placebo.

1 = Udca.

```
tratamiento
Placebo   UDCA
      84      85
```

- Por tratamiento y estatus

```
tratamiento
estatus  0  1
      0 39 58
      1 45 27
```

- Por etapa

0= etapa 1

1= etapa 2

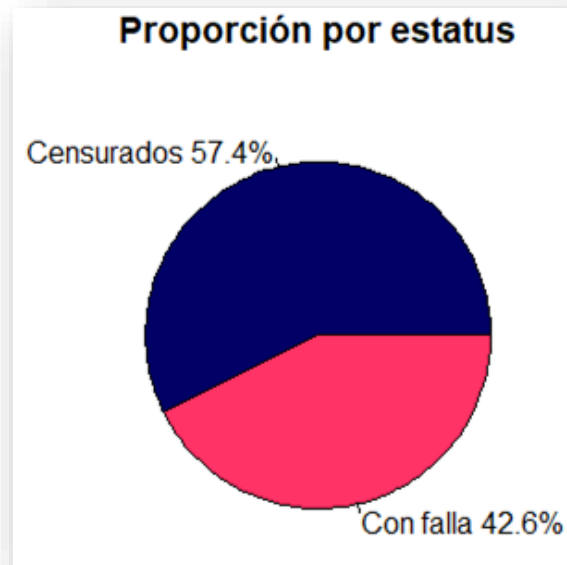
```
etapa
etapa 1 etapa 2
      52      117
```

- Por etapa y estatus

```
Etapa
Estatus  0  1
      0 34 63
      1 18 54
```

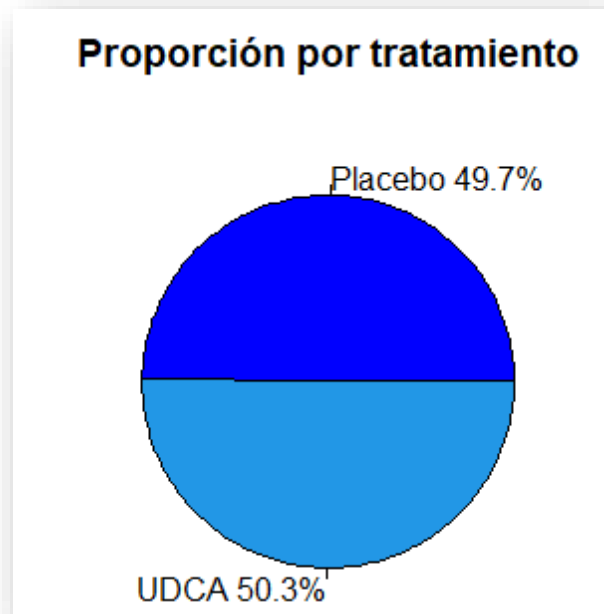
## Gráficos

- **Estatus**



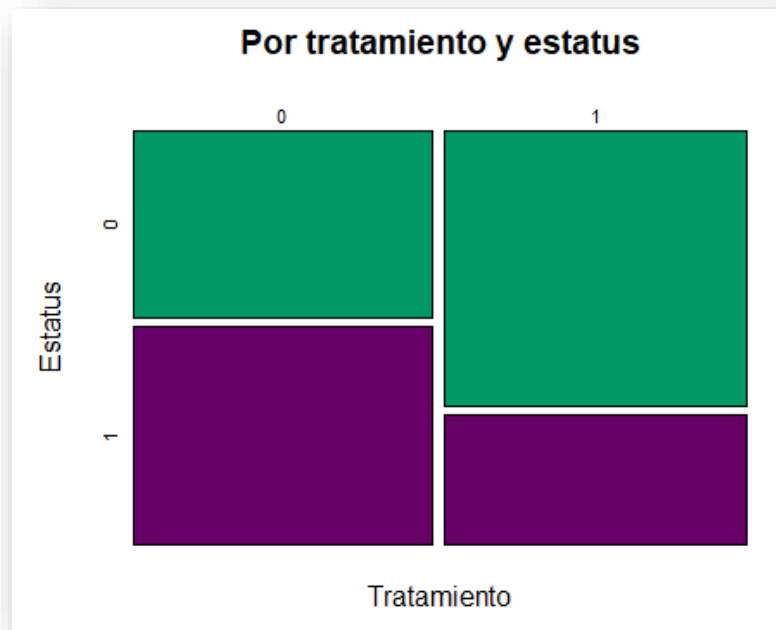
En el gráfico se observa que el 57.4% de los pacientes son censurados, mientras que el 42.6% muestran alguna falla.

- **Tratamiento**



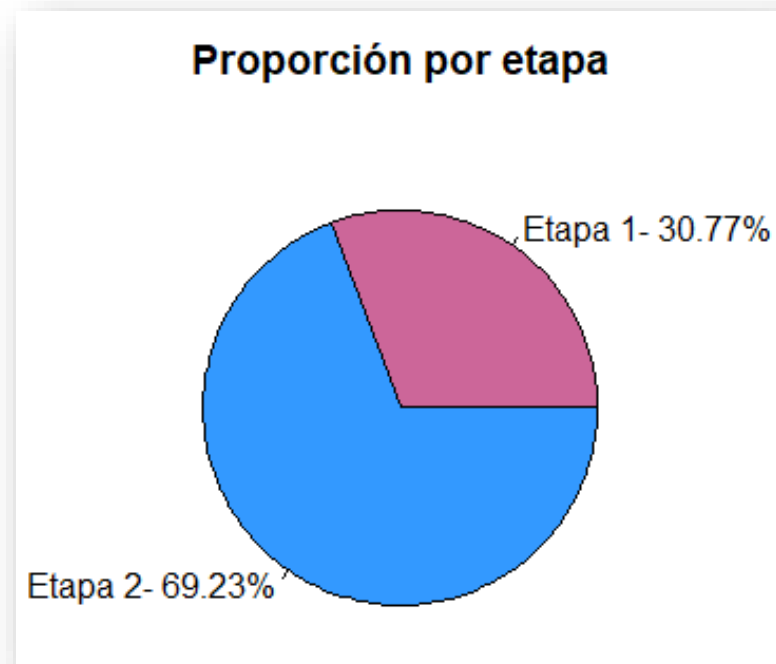
Podemos observar que la aplicación de los tratamientos fue equitativa.



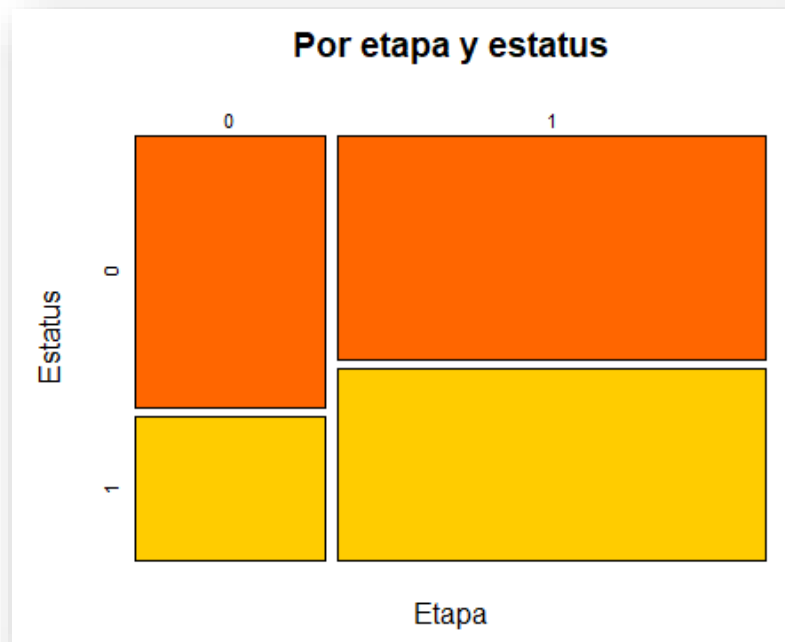


Se observa en el gráfico que el estatus de censura es mayor para aquellos que llevaron el tratamiento de Udca.

- **Etapas**

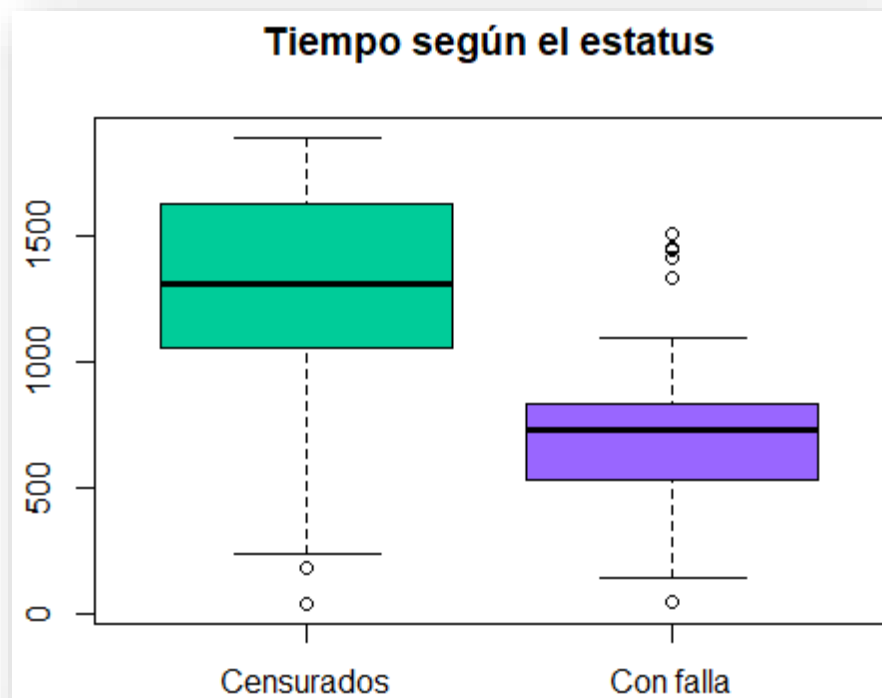


Casi el 70% de los pacientes están en la etapa 2, mientras solo un poco más del 30% son parte de la etapa 1.



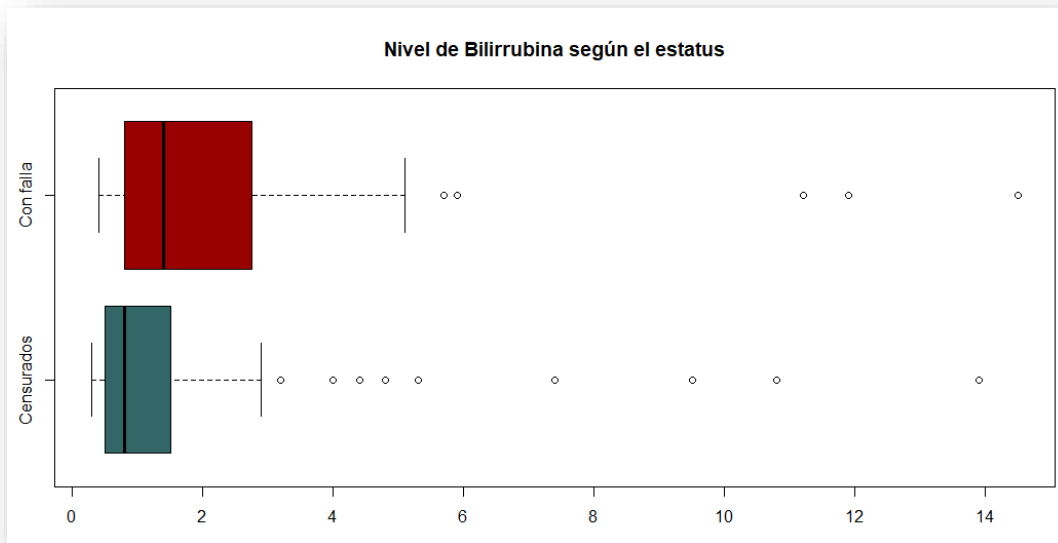
La mayor cantidad de pacientes censurados se ubican en la etapa 2.

- **Tiempo**



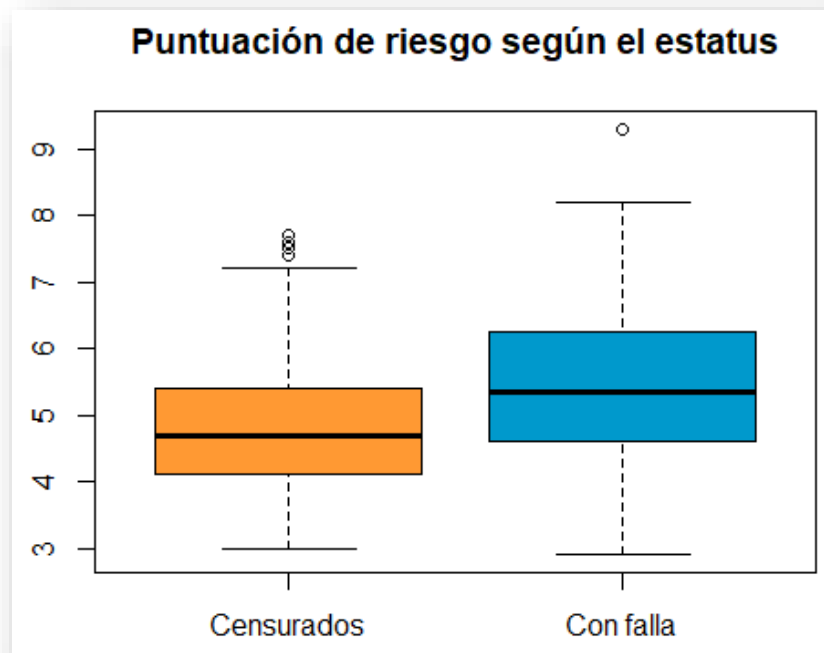
El tiempo registrado hasta una falla en el hígado del paciente es menor al tiempo registrado de los pacientes censurados.

- **Bilirrubina**



Los pacientes registrados con alguna falla en el hígado poseen un nivel más alto de bilirrubina de aquellos que pacientes censurados.

- **Puntuación de riesgo**



La puntuación de riesgo de mayo es mayor para aquellos pacientes registrados con alguna falla en el hígado.

## Análisis de regresión para selección de variables

### Modelo de regresión

#### Modelo con todas las variables

##### Prueba de significancia de cada variable para la regresión

**H0:** La variable no es significativa para el modelo

**Ha:** La variable es significativa para el modelo

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.4579294	1.1848208	-0.386	0.69913
udca1\$futime	-0.0030495	0.0005418	-5.628	1.82e-08 ***
udca1\$trt	-0.2954990	0.4063177	-0.727	0.46707
udca1\$stage	0.1670474	0.4334694	0.385	0.69996
udca1\$bili	-0.2148857	0.1188711	-1.808	0.07065 .
udca1\$riskscore	0.7073148	0.2579847	2.742	0.00611 **

Con una confianza de 90%, podemos observar que la variable “status” posee el  $p$ -valor más alto y dado que  $p$ -valor  $> \alpha$ , no es significativa. Por lo tanto, se eliminará del modelo.

#### Modelo sin la variable “stage”

##### Prueba de significancia de cada variable para la regresión

**H0:** La variable no es significativa para el modelo

**Ha:** La variable es significativa para el modelo

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.4434105	1.1790137	-0.376	0.70685
udca1\$futime	-0.0030468	0.0005416	-5.626	1.85e-08 ***
udca1\$trt	-0.3100582	0.4040576	-0.767	0.44287
udca1\$bili	-0.2119566	0.1182345	-1.793	0.07302 .
udca1\$riskscore	0.7259504	0.2527301	2.872	0.00407 **

Con una confianza de 90%, podemos observar que, en el ajuste del modelo, ahora la variable “trt” posee el  $p$ -valor más alto y dado que  $p$ -valor  $> \alpha$ , no es significativa. Por lo tanto, se eliminará del modelo.

#### Modelo sin la variable “stage” y “trt”

##### Prueba de significancia de cada variable para la regresión

**H0:** La variable no es significativa para el modelo

**Ha:** La variable es significativa para el modelo

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.4586607	1.1712295	-0.392	0.69535
udca1\$futime	-0.0031743	0.0005238	-6.060	1.36e-09 ***
udca1\$bili	-0.2194551	0.1176075	-1.866	0.06204 .
udca1\$riskscore	0.7262256	0.2520722	2.881	0.00396 **

Con una confianza de 90%, todas las variables son significativas para el modelo ya que el  $p\text{-valor} < \alpha$  en cada una de las variables. Por lo tanto, este será el modelo seleccionado.

La variable con mayor significancia es “futime” siendo esta el tiempo.

La variable menos significativa es “bili” representando el nivel de bilirrubina del paciente.

### Coeficientes del modelo

(Intercept)

0.632129693841599

udca1\$futime

0.996830705015335

udca1\$bili

0.802956241068207

udca1\$riskscore

2.06726320356718

Hay una asociación positiva ente la puntuación de riesgo y el estatus, es decir, a mayor puntuación de riesgo, mayor es la probabilidad de que el hígado se mantenga funcionando. Por otro lado, a menor tiempo y bilirrubina, mayor es la probabilidad de que el hígado continúe funcionando.

### Evaluación del modelo

El análisis mediante la función ANOVA ayuda a comprobar la validez del modelo.

#### Prueba de significancia para el modelo de regresión

**H0:** El modelo no es significativo

**Ha:** El modelo es significativo

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
	<int>	<dbl>	<int>	<dbl>	<dbl>
<b>NULL</b>	NA	NA	168	230.5719	NA
<b>udca1\$futime</b>	1	56.9154302	167	173.6565	4.549620e-14
<b>udca1\$bili</b>	1	0.1786797	166	173.4778	6.725103e-01
<b>udca1\$riskscore</b>	1	9.0533625	165	164.4244	2.622121e-03

Con un nivel de confianza del 90%, se rechaza  $H_0$ , por lo que concluimos que el modelo seleccionado con las variables “futime”, “bili” y “riskscore” es significativo.

### Función de la regresión.

$$P(Y) = \frac{e^{0.632+0.997 \text{ futime} +0.803 \text{ bili} +2.067 \text{ riskscore}}}{1 + e^{0.632+0.997 \text{ futime} +0.803 \text{ bili} +2.067 \text{ riskscore}}}$$

Análisis de supervivencia de forma

### Modelo de Cox

#### Modelo con todas las variables

##### Prueba de significancia del modelo Cox

**H0:** Ninguna variable es significativa en el modelo

**Ha:** Al menos una variable es significativa

Concordance= 0.682 (se = 0.032 )  
 Likelihood ratio test= 30.34 on 5 df, p=1e-05  
 Wald test = 29.86 on 5 df, p=2e-05  
 Score (logrank) test = 31.5 on 5 df, p=7e-06

Con una confianza al 95% tenemos que,  $p\text{-valor} < \alpha$ . Por lo tanto, se rechaza  $H_0$ , al menos una variable del modelo es significativa.

	coef	exp(coef)	se(coef)	z	Pr(> z )	
u1\$trt	-1.017987	0.361322	0.256614	-3.967	7.28e-05	***
u1\$id	0.000358	1.000358	0.002612	0.137	0.8910	
u1\$stage	0.041310	1.042175	0.292069	0.141	0.8875	
u1\$bili	0.089380	1.093496	0.078331	1.141	0.2538	
u1\$riskscore	0.310069	1.363519	0.161706	1.917	0.0552	.

Teniendo un nivel de confianza del 90%, podemos observar que la variable “id” no es significativa para el modelo, ya que su  $p\text{-valor} > \alpha$ . Por lo tanto, se eliminará del modelo.

##### Prueba de significancia del modelo Cox

**H0:** Ninguna variable es significativa en el modelo

**Ha:** Al menos una variable es significativa

Concordance= 0.684 (se = 0.031 )  
 Likelihood ratio test= 30.32 on 4 df, p=4e-06  
 Wald test = 29.86 on 4 df, p=5e-06  
 Score (logrank) test = 31.47 on 4 df, p=2e-06

Con un nivel de confianza del 90% tenemos que,  $p\text{-valor} < \alpha$ . Por lo tanto, se rechaza  $H_0$ , al menos una variable del modelo es significativa.

	coef	exp(coef)	se(coef)	z	Pr(> z )	
udca1.new\$strtr	-1.01636	0.36191	0.25628	-3.966	7.31e-05	***
udca1.new\$stage	0.03453	1.03514	0.28768	0.120	0.9045	
udca1.new\$bili	0.08971	1.09385	0.07827	1.146	0.2518	
udca1.new\$riskscore	0.30950	1.36275	0.16142	1.917	0.0552	.

Teniendo una confianza al 90%, podemos observar que la variable “stage” no es significativa para el modelo, ya que su  $p\text{-valor} > \alpha$ . Por lo tanto, se eliminará del modelo.

### Prueba de significancia del modelo Cox

**H0:** Ninguna variable es significativa en el modelo

**Ha:** Al menos una variable es significativa

Concordance= 0.684 (se = 0.032 )  
 Likelihood ratio test= 30.3 on 3 df, p=1e-06  
 Wald test = 29.83 on 3 df, p=1e-06  
 Score (logrank) test = 31.46 on 3 df, p=7e-07

Con una confianza al 90% tenemos que,  $p\text{-valor} < \alpha$ . Por lo tanto, se rechaza  $H_0$ , al menos una variable del modelo es significativa.

	coef	exp(coef)	se(coef)	z	Pr(> z )	
udca1.new\$strtr	-1.01972	0.36069	0.25485	-4.001	6.3e-05	***
udca1.new\$bili	0.08913	1.09322	0.07790	1.144	0.2526	
udca1.new\$riskscore	0.31446	1.36952	0.15594	2.017	0.0437	*

Teniendo una confianza al 90%, podemos observar que la variable “bili” no es significativa para el modelo, ya que su  $p\text{-valor} > \alpha$ . Por lo tanto, se eliminará del modelo.

### Prueba de significancia del modelo Cox

**H0:** Ninguna variable es significativa en el modelo

**Ha:** Al menos una variable es significativa

Concordance= 0.681 (se = 0.032 )  
 Likelihood ratio test= 29.11 on 2 df, p=5e-07  
 Wald test = 28.46 on 2 df, p=7e-07  
 Score (logrank) test = 29.46 on 2 df, p=4e-07

Con una confianza al 90% tenemos que,  $p\text{-valor} < \alpha$ . Por lo tanto, se rechaza  $H_0$ , al menos una variable del modelo es significativa.

	coef	exp(coef)	se(coef)	z	Pr(> z )	
udca1.new\$strtr	-0.9656	0.3808	0.2473	-3.904	9.47e-05	***
udca1.new\$riskscore	0.4421	1.5560	0.1069	4.136	3.53e-05	***

Teniendo una confianza al 90%, podemos observar que las variables “trt” y “riskscore” son significativas ya que su  $p\text{-valor} < \alpha$ .

Obteniendo así un modelo de Cox significativo con las variables de tratamiento y de puntaje de riesgo, esto al 90% de confianza. Los valores del exp(coef) generados por el modelo seleccionado son los siguientes:

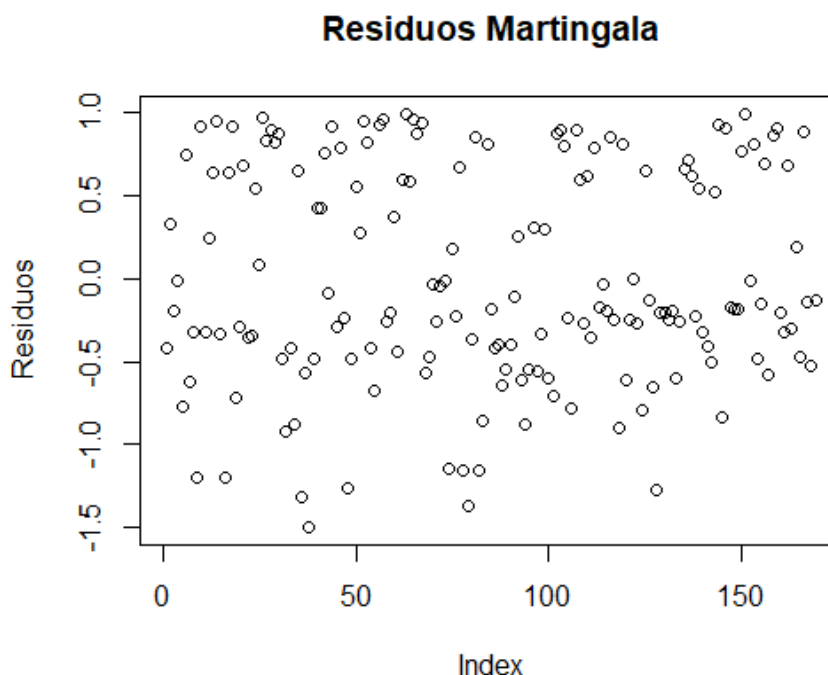
	exp(coef)	exp(-coef)	lower .95	upper .95
u1\$strtr	0.3808	2.6263	0.2345	0.6183
u1\$riskscore	1.5560	0.6427	1.2619	1.9186

Donde:

- Tratamiento es una variable discreta la cual nos indica, donde se puede apreciar que aquellos pacientes que tomaron placebo tienen 0.3808 mayor riesgo que los que tomaron el Udca de sufrir una falla o deterioro del mantenimiento del hígado.
- Riskscore es una variable continua que indica que el riesgo por unidad del puntaje de riesgo de mayo posee un incremento del 1.5560.

## Análisis de residuales

- **Residuos Martingala**

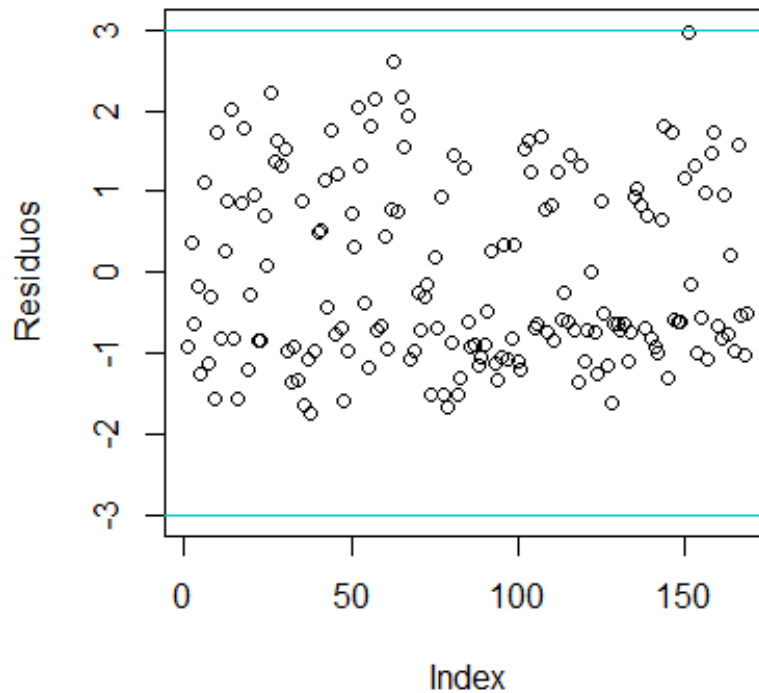


Se puede observar que los datos se distribuyen en valores de  $-\infty$  a 1, siendo este un buen indicador para los residuos de martingala,



- **Desviación Residual**

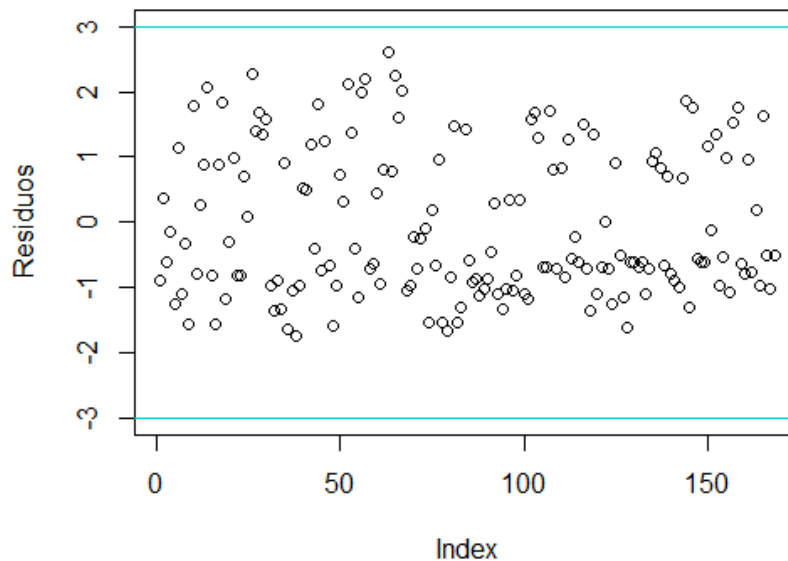
### Residuos de Desvio



**Valor atípico:**

**151 2.964**

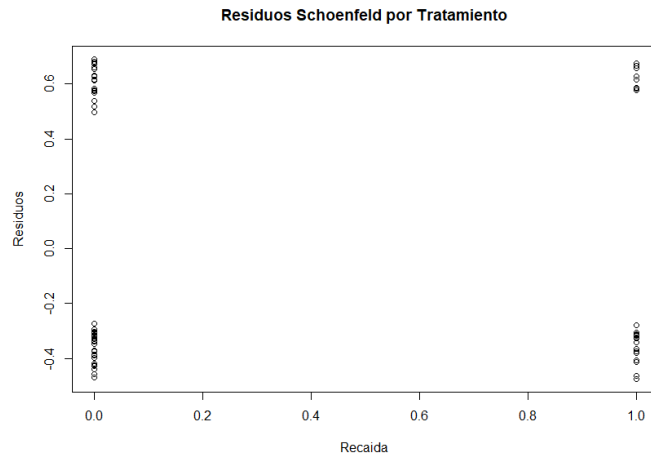
### Residuos de Desvio



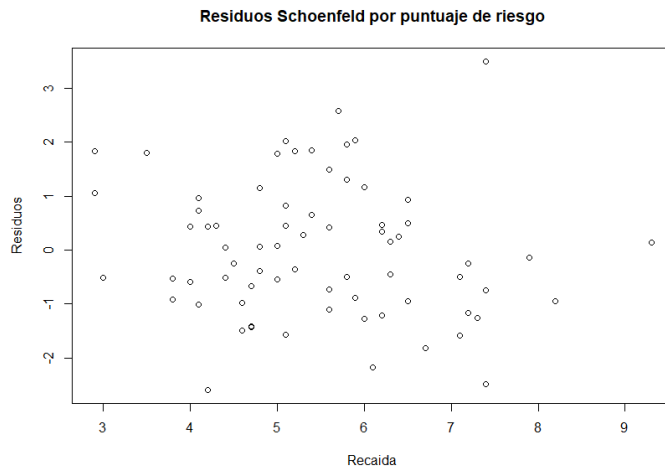
Gráfica sin atípico:

Se puede observar que los datos se distribuyen en valores de -3 a 3, mostrando un buen comportamiento de los datos sin algún valor atípico.

- **Residuales de Schoenfeld**

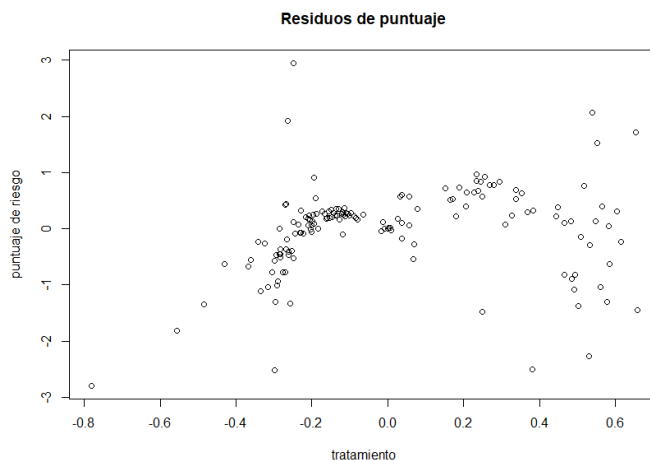


Al ser una variable discreta, se puede observar que los datos muestran la separación por categoría. Por lo tanto, es un buen comportamiento de los datos.



Al ser una variable continua, los datos no muestran ningún patrón ni dispersión, lo cual indica un buen comportamiento de los datos.

- **Residuales de Puntaje**



Se puede observar que los datos muestran dispersión por lo tanto es un buen comportamiento de los datos.

En conclusión, en los gráficos se puede observar que en todos los residuales se muestra una dispersión y proporción adecuados.

## **Comparación del modelo Cox y la regresión logística**

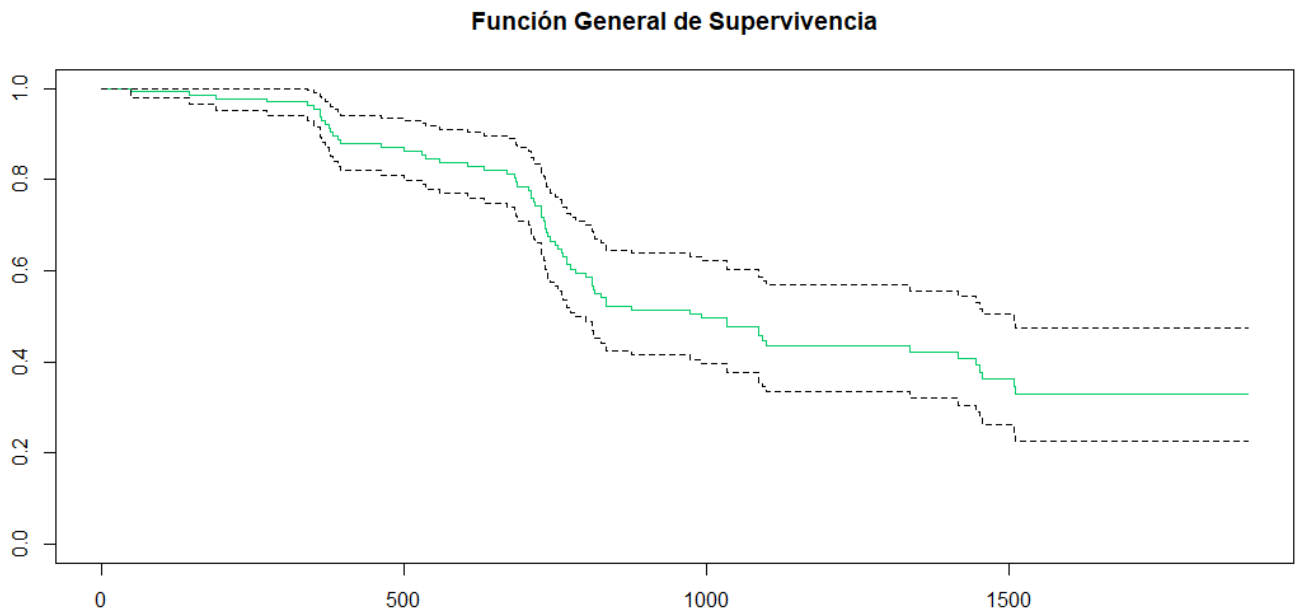
En comparación con el modelo de regresión logística y el modelo Cox obtenido, podemos observar que en la regresión logística se obtuvieron como variables significativas el tiempo, los niveles de bilirrubina, así como el puntaje de riesgo de mayo. Por el otro lado, en el modelo de Cox se obtuvieron como variables significativas únicamente al tratamiento y el puntaje de riesgo de mayo.

Ambos modelos mostraron la importancia de la variable de puntaje de riesgo para el modelo.

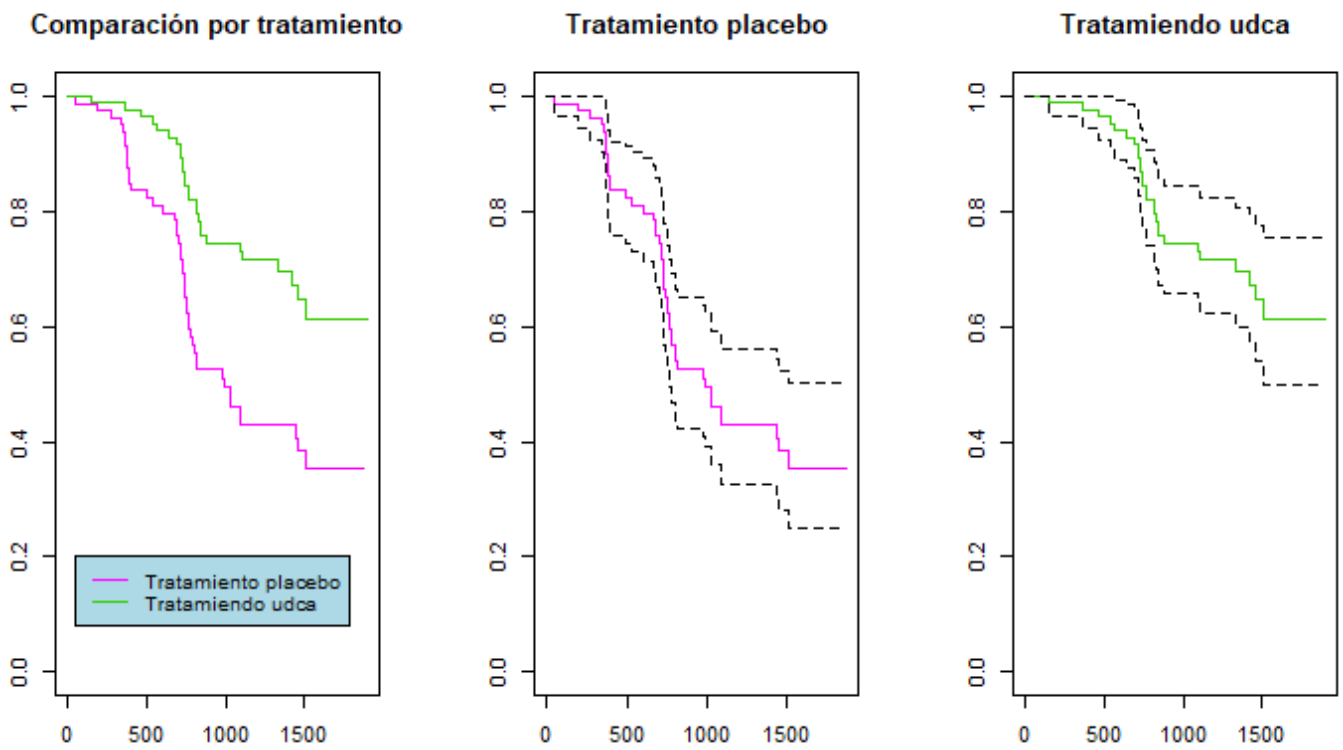
Es fundamental observar que mientras que para la regresión el tratamiento es indiferente pero los niveles de bilirrubina son significantes, para el modelo Cox es precisamente lo contrario para estas 2 variables.

## Método Kaplan-Meier

- **Función de Supervivencia General**



- **Estimación de modelo de Kaplan Meier por tratamiento**



## Prueba de igualdad de funciones de supervivencia

**H<sub>0</sub>:** Las funciones de supervivencia son iguales

**H<sub>a</sub>:** Las funciones difieren

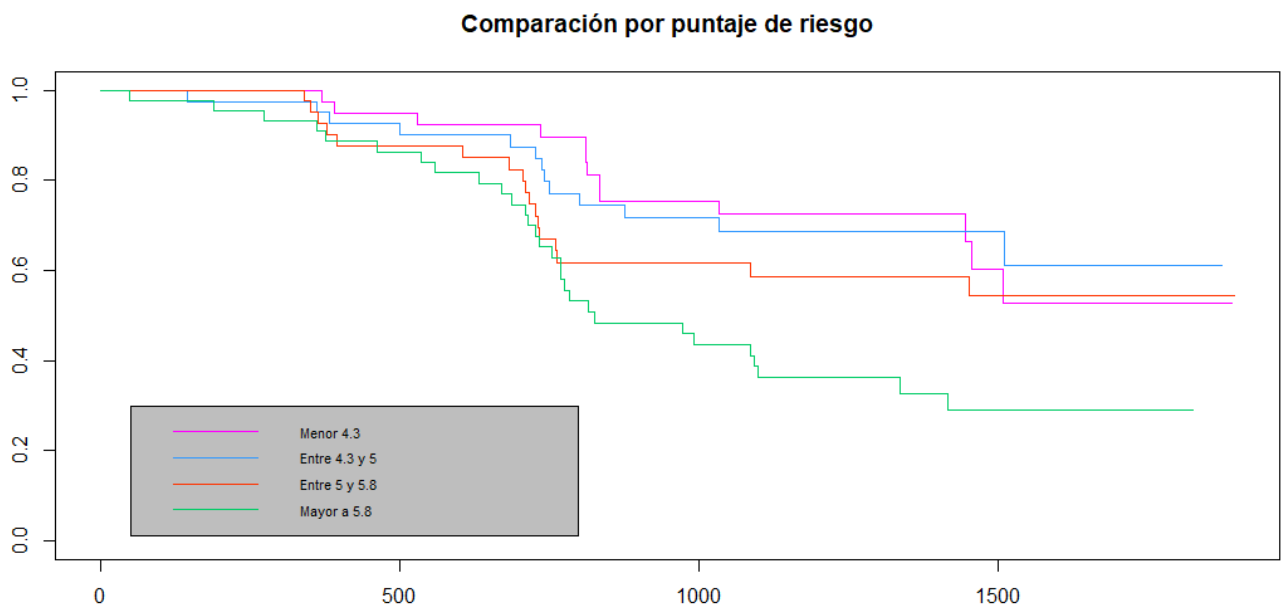
	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
u1\$trt=0	84	45	29.9	7.68	13.2
u1\$trt=1	85	27	42.1	5.44	13.2

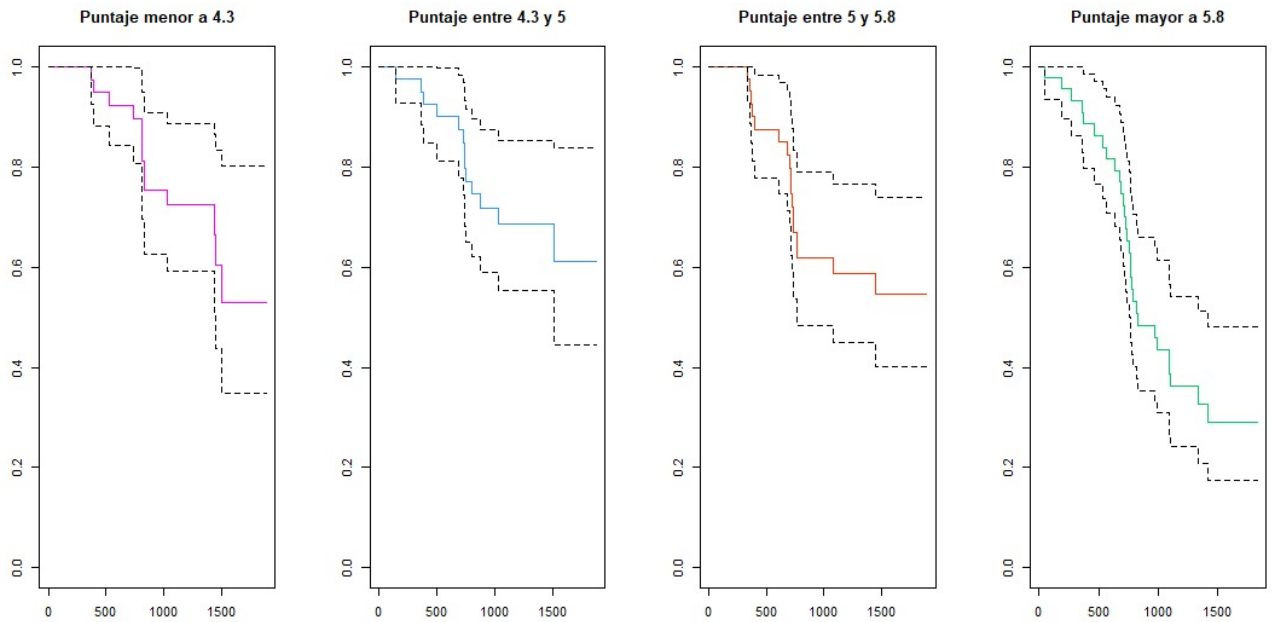
Chisq= 13.2 on 1 degrees of freedom, p= 3e-04

Con un 90% de confianza tenemos que dado que el ***p*-valor** <  $\alpha$ , SI se rechaza H<sub>0</sub>, por lo tanto, las funciones difieren.

En conclusión, se comprobó analíticamente que los tratamientos dados difieren en la función de supervivencia. Gráficamente, podremos concluir que el tratamiento Udca nos brinda una mayor supervivencia al hígado.

- **Estimación de modelo de Kaplan Meier por puntaje de riesgo**



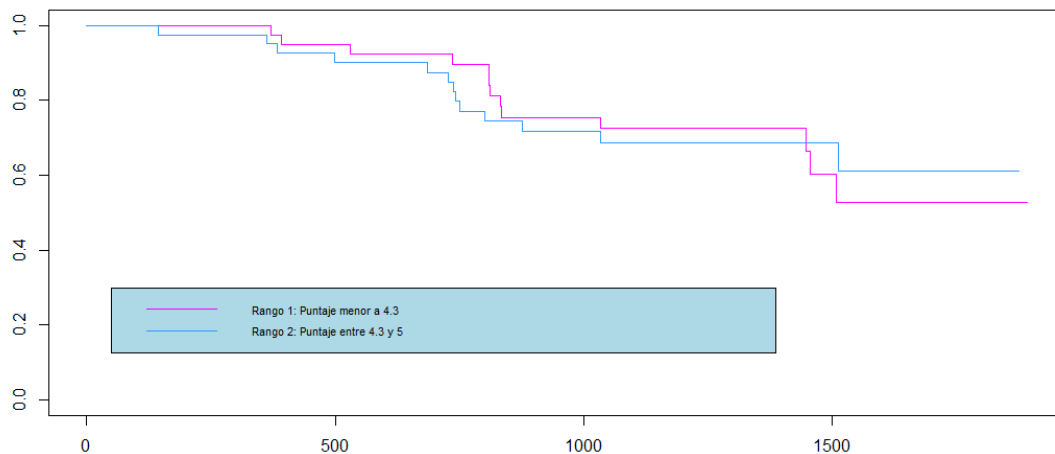


### Prueba de igualdad de funciones de supervivencia

**H0:** Las funciones de supervivencia son iguales

**Ha:** Las funciones difieren

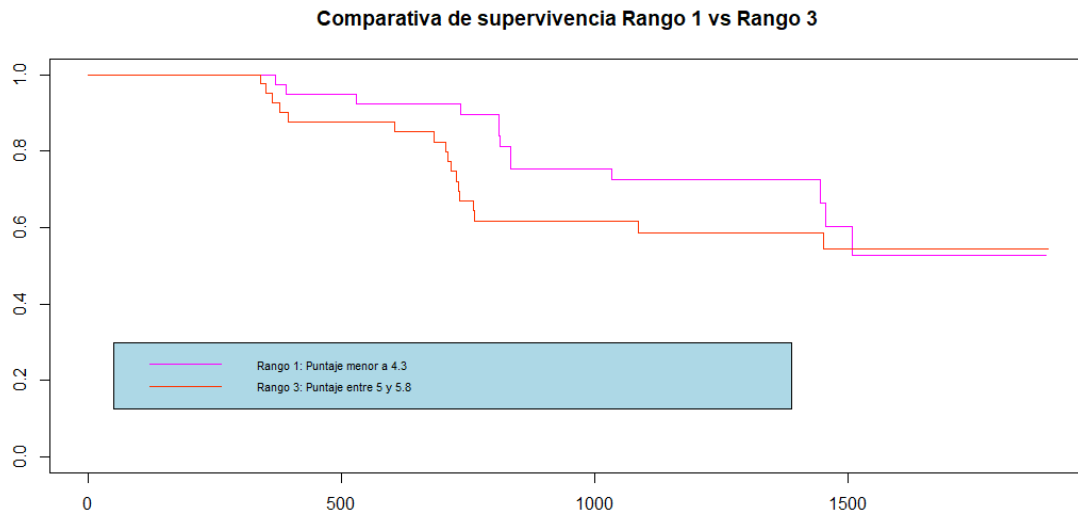
Comparativa de supervivencia Rango 1 vs Rango 2



	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
cri[cri != "3" & cri != "4"]=1	42	13	13.1	0.000302	0.000608
cri[cri != "3" & cri != "4"]=2	40	13	12.9	0.000305	0.000608

Chisq= 0 on 1 degrees of freedom, p= 1

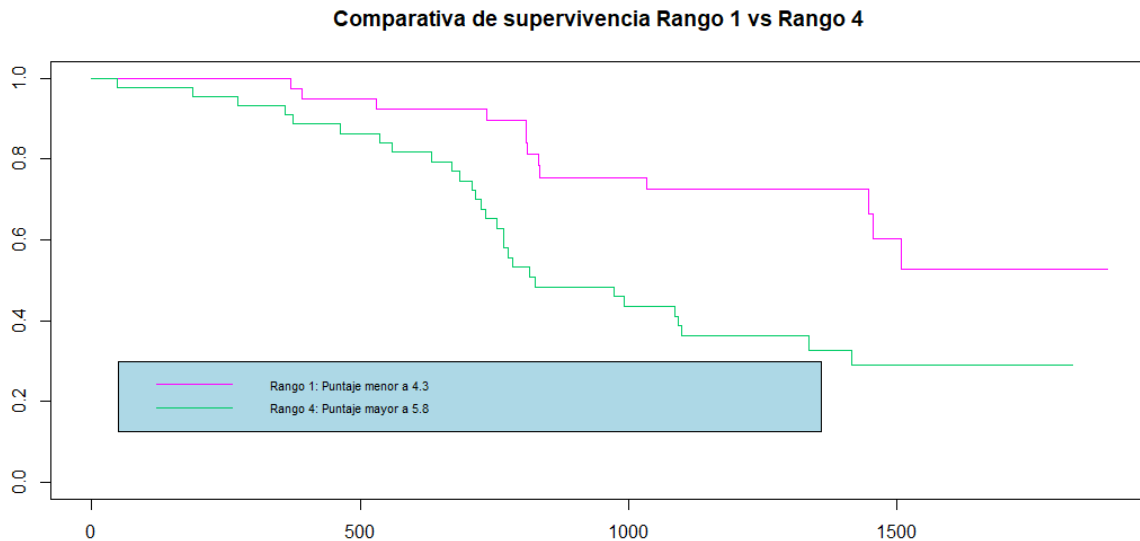
Con un 90% de confianza, tenemos que  $p\text{-valor} > \alpha$ , No se rechaza H0, por lo tanto, las funciones son iguales, es decir, un puntaje de riesgo menor a 4.3 y un puntaje entre 4.3 y 5 no difieren de manera significativa.



	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
cri[cri != "2" & cri != "4"]=1	42	13	15.7	0.451	0.95
cri[cri != "2" & cri != "4"]=3	42	17	14.3	0.492	0.95

Chisq= 1 on 1 degrees of freedom, p= 0.3

Con un 90% de confianza, tenemos que  $p\text{-valor} > \alpha$ , No se rechaza  $H_0$ , por lo tanto, las funciones son iguales, es decir, un puntaje de riesgo menor a 4.3 y un puntaje entre 5 y 5.8 no difieren de manera significativa.

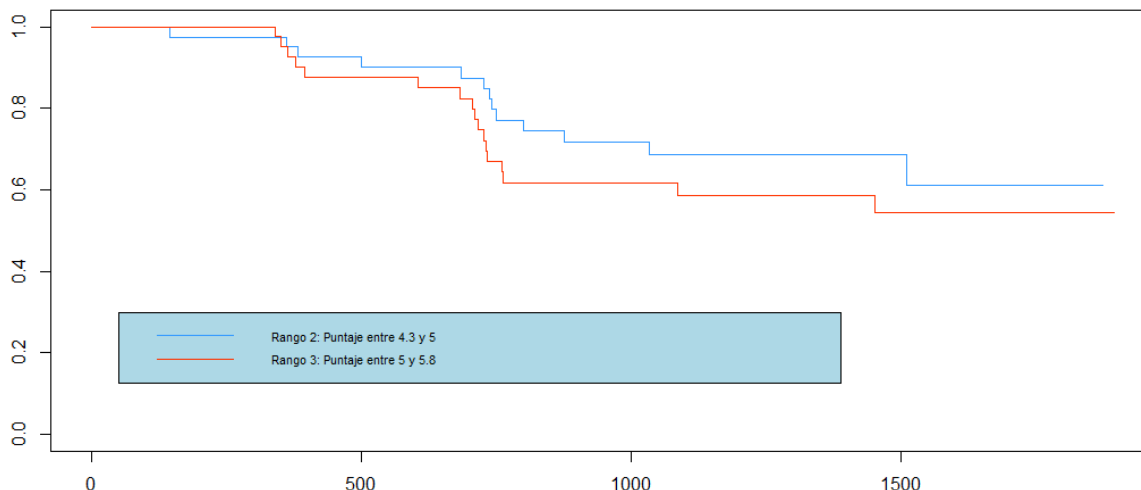


	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
cri[cri != "2" & cri != "3"]=1	42	13	22.8	4.24	9.38
cri[cri != "2" & cri != "3"]=4	45	29	19.2	5.06	9.38

Chisq= 9.4 on 1 degrees of freedom, p= 0.002

Con un 90% de confianza, tenemos que  $p\text{-valor} < \alpha$ , SI se rechaza  $H_0$ , por lo tanto, las funciones difieren, es decir, un puntaje de riesgo menor a 4.3 y un puntaje mayor a 5 difieren de manera significativa.

Comparativa de supervivencia Rango 2 vs Rango 3

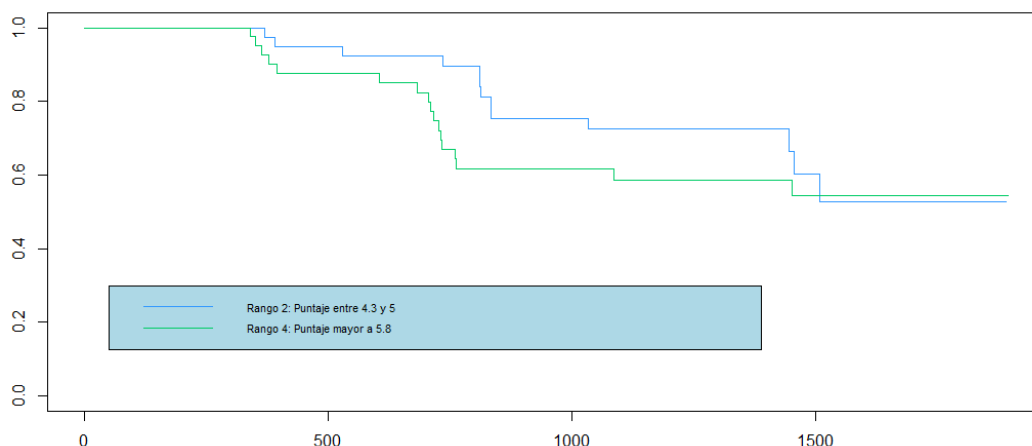


	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
cri[cri != "1" & cri != "4"]=2	40	13	15.6	0.432	0.904
cri[cri != "1" & cri != "4"]=3	42	17	14.4	0.468	0.904

Chisq= 0.9 on 1 degrees of freedom, p= 0.3

Con un 90% de confianza, tenemos que  $p\text{-valor} > \alpha$ , No se rechaza  $H_0$ , por lo tanto, las funciones son iguales, es decir, un puntaje de riesgo entre 4.3 y 5 y un puntaje entre 5 y 5.8 no difieren de manera significativa.

Comparativa de supervivencia Rango 2 vs Rango 4



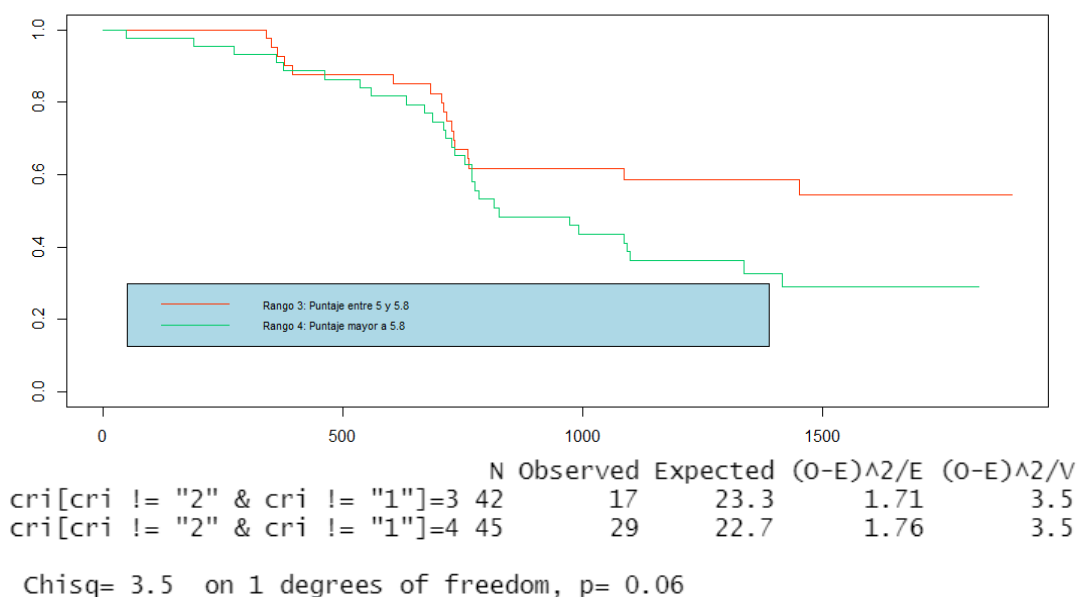
	N	Observed	Expected	(O-E) <sup>2</sup> /E	(O-E) <sup>2</sup> /V
cri[cri != "1" & cri != "3"]=2	40	13	22.3	3.88	8.32
cri[cri != "1" & cri != "3"]=4	45	29	19.7	4.39	8.32

Chisq= 8.3 on 1 degrees of freedom, p= 0.004



Con un 90% de confianza, tenemos que  $p\text{-valor} > \alpha$ , No se rechaza  $H_0$ , por lo tanto, las funciones son iguales, es decir, un puntaje de riesgo entre 4.3 y 5 y un puntaje mayor a 5.8 no difieren de manera significativa.

Comparativa de supervivencia Rango 3 vs Rango 4



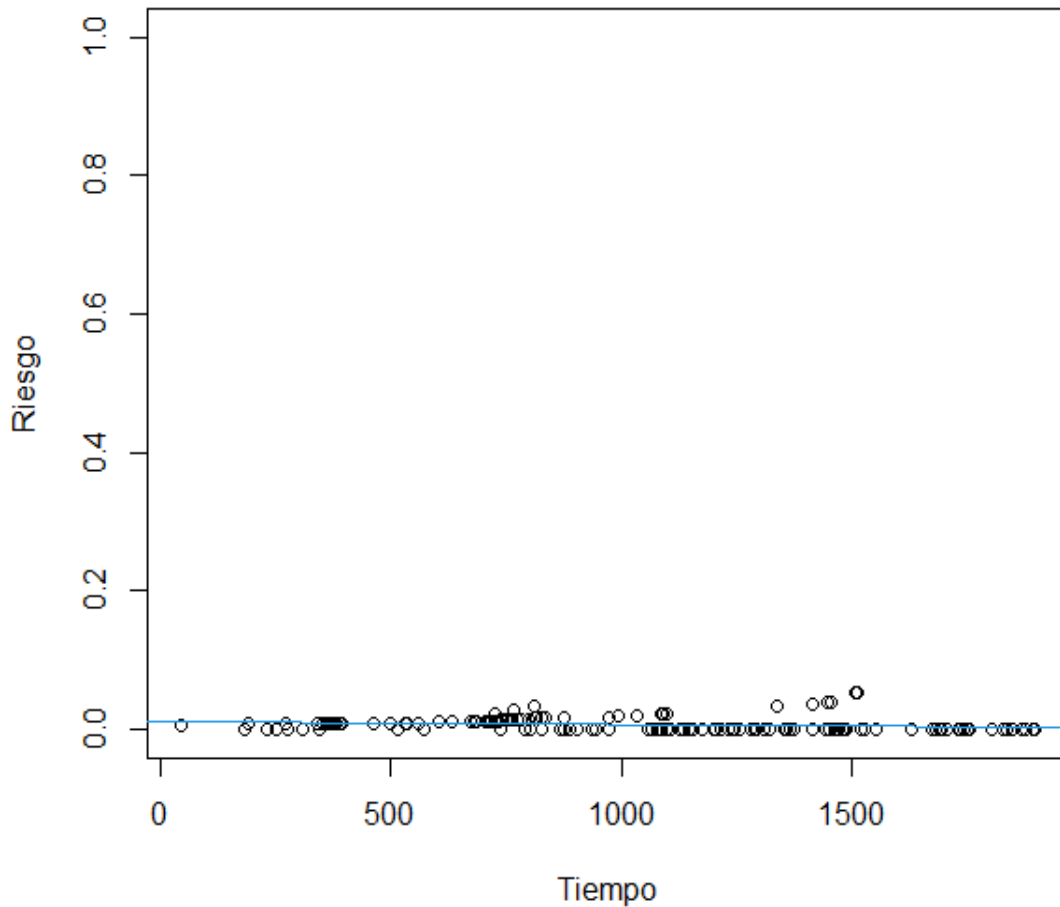
Con un 90% de confianza, tenemos que  $p\text{-valor} < \alpha$ , SI se rechaza  $H_0$ , por lo tanto, las funciones difieren, es decir, un puntaje de riesgo entre 5 y 5.8 y un puntaje mayor a 5.8 difieren de manera significativa.

En conclusión, se comprobó analíticamente que un puntaje de riesgo de Mayo PBC mayor a 5.8 difieren de manera significativa. Gráficamente, podremos concluir que al tener un porcentaje arriba de 5.8 se tiene una menor supervivencia del hígado.

## Modelos paramétricos

- Familia Exponencial

### Funcion de Riesgo



Identificar si en la gráfica de la función de riesgo estimada contra el tiempo, se observa una línea recta horizontal.

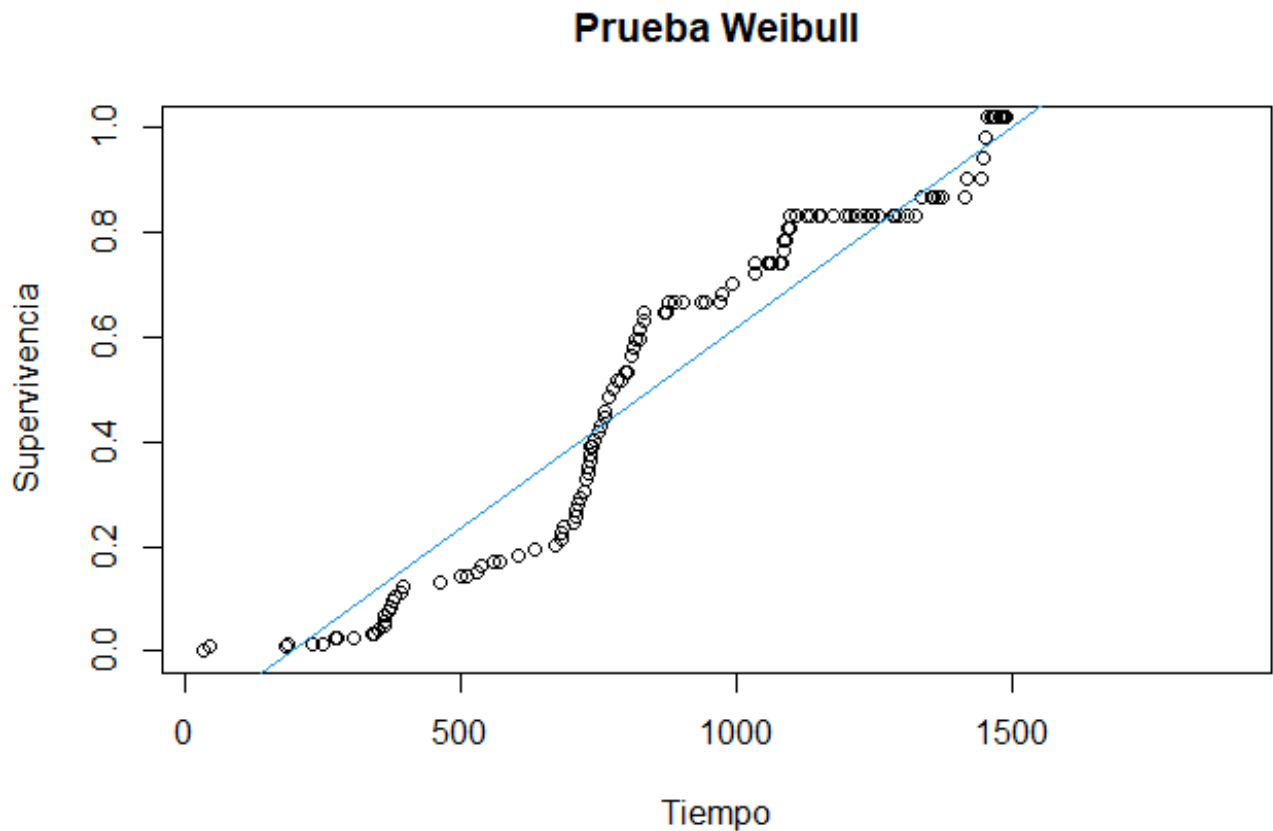
**H<sub>0</sub>:** No muestra una tendencia lineal.

**H<sub>a</sub>:** Muestra tendencia lineal.

Residual standard error: 0.01062 on 154 degrees of freedom  
Multiple R-squared: 0.03187, Adjusted R-squared: 0.02559  
F-statistic: 5.07 on 1 and 154 DF, p-value: 0.02576

Con un 90% de confianza,  $p\text{-valor} < \alpha$ , rechazamos H<sub>0</sub>, la gráfica muestra una tendencia lineal, sin embargo, dado que tenemos una R ajustada del 2.57%, el modelo NO sigue una recta horizontal, por lo tanto, no sugiere un modelo exponencial.

- Familia Weibull



Identificar si en la gráfica del -logaritmo de la función de supervivencia estimada contra el tiempo, se observa una línea recta que corta el origen.

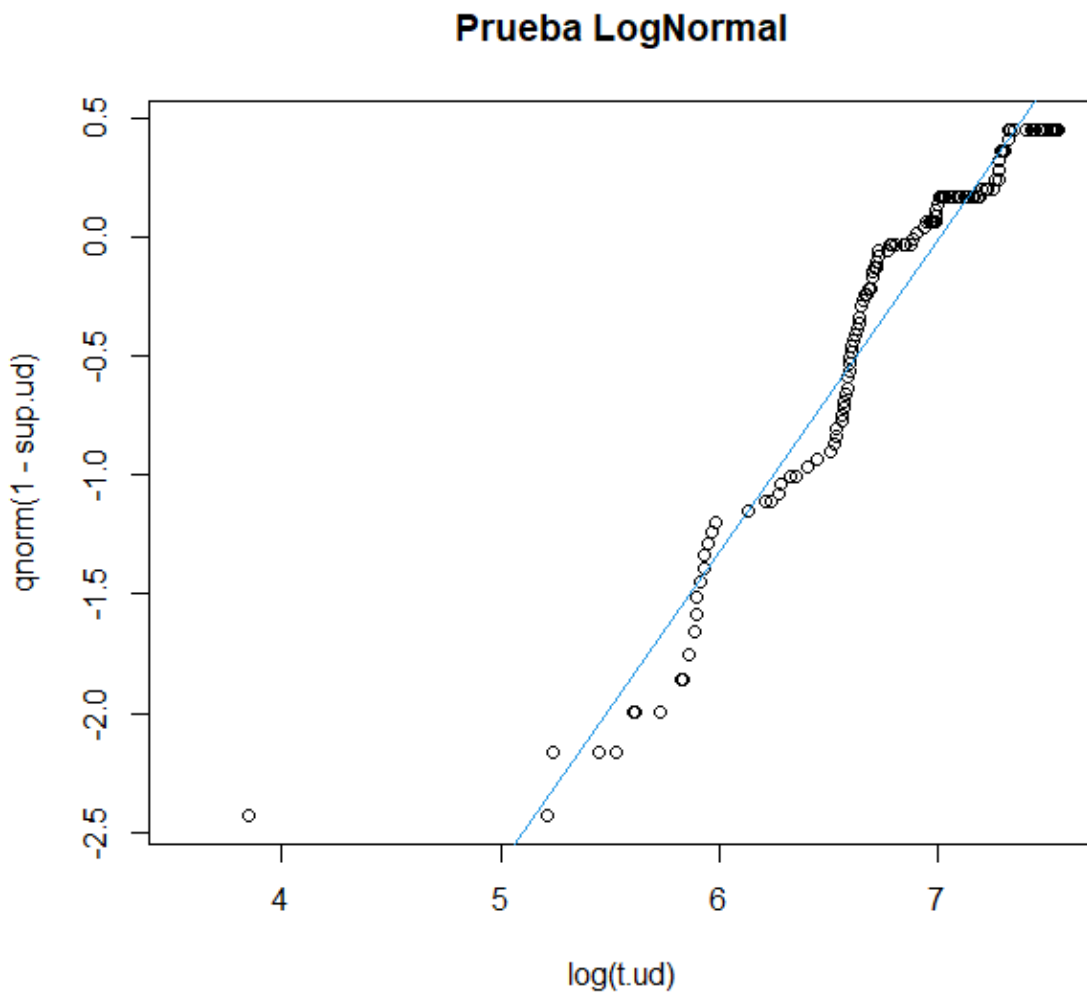
**H<sub>0</sub>:** No muestra una tendencia lineal.

**H<sub>a</sub>:** Muestra tendencia lineal.

Residual standard error: 0.08765 on 155 degrees of freedom  
 Multiple R-squared: 0.9453, Adjusted R-squared: 0.945  
 F-statistic: 2680 on 1 and 155 DF, p-value: < 2.2e-16

Con un 90% de confianza,  **$p\text{-valor} < \alpha$** , rechazamos H<sub>0</sub>, la gráfica muestra una tendencia lineal, sin embargo, dado que la recta no cruza por el origen, la prueba NO sugiere un modelo Weibull.

- Familia LogNormal



Identificar si en la gráfica de la función de riesgo estimada contra el tiempo, se observa una línea recta horizontal.

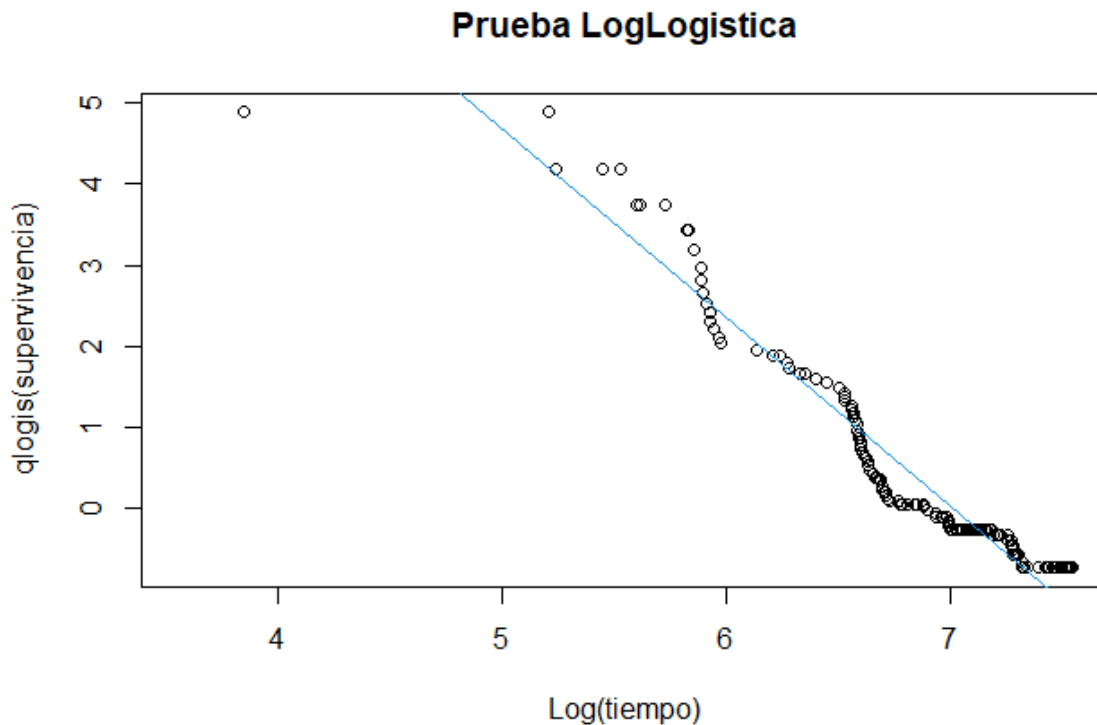
**H<sub>0</sub>:** No muestra una tendencia lineal.

**H<sub>a</sub>:** Muestra tendencia lineal.

Residual standard error: 0.1496 on 153 degrees of freedom  
 Multiple R-squared: 0.957, Adjusted R-squared: 0.9567  
 F-statistic: 3402 on 1 and 153 DF, p-value: < 2.2e-16

Con un 90% de confianza, ***p-valor*** < ***α***, rechazamos H<sub>0</sub>, la gráfica muestra una tendencia lineal. Además, tenemos una R ajustada del 95.67%, podemos concluir que el modelo sigue una recta, por lo tanto, se sugiere que el modelo sea LogNormal.

- Familia LogLogistica



Identificar si en la gráfica del  $\text{Logit}S(t)$  contra logaritmo del tiempo, se observa una línea recta.  $S(t)$  es la función de supervivencia estimada.

**H<sub>0</sub>:** No muestra una tendencia lineal.

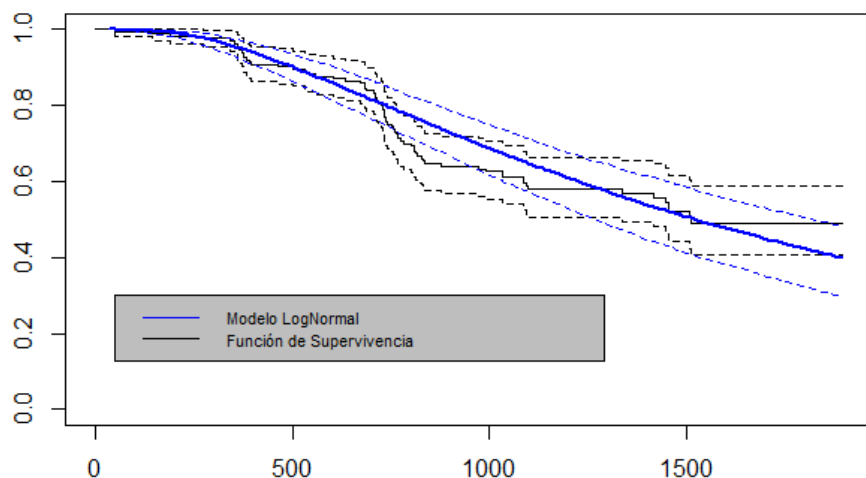
**H<sub>a</sub>:** Muestra tendencia lineal.

Residual standard error: 0.301 on 153 degrees of freedom  
 Multiple R-squared: 0.9454, Adjusted R-squared: 0.9451  
 F-statistic: 2650 on 1 and 153 DF, p-value:  $< 2.2e-16$

Con un 90% de confianza,  $p\text{-valor} < \alpha$ , rechazamos  $H_0$ , la gráfica muestra una tendencia lineal. Además, tenemos una R ajustada del 94.51%, podemos concluir que el modelo sigue una recta, por lo tanto, se sugiere que el modelo sea LogLogístico.

## Comparación de modelos

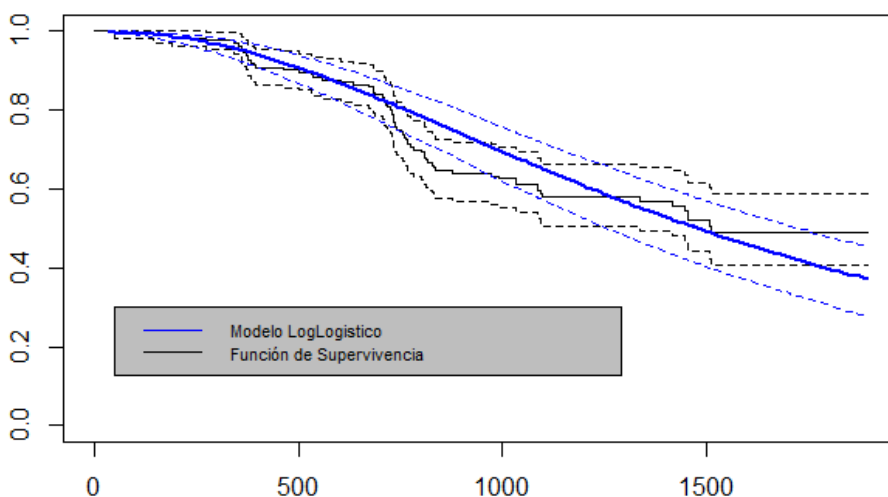
LogNormal



**AIC:** 1227.632

**Ajuste:** 95.67%

LogLogistico



**AIC:** 1226.121

**Ajuste:** 94.51%

Considerando el ajuste de ambos modelos paramétricos, podemos deducir que el modelo LogNormal se adecua mejor a los datos de supervivencia originales.

### Modelo LogNormal

El modelo LogNormal se constituye de 2 parámetros,  $\mu, \sigma$ , donde:

$$\mu = -9.185901$$

$$\sigma = 1.310402$$

**Función de riesgo:**  $\frac{f(t)}{S(t)}$

**Función de densidad:**

**Función de Supervivencia:**

$$S(t) = 1 - \Phi\left(\frac{\ln(t) - (-9.1859)}{1.3104}\right)$$

$$f(t) = \frac{\phi\left(\frac{\ln(t) - (-9.1859)}{1.3104}\right)}{t}$$

## Conclusiones

Gracias al trabajo realizado, logramos determinar cuál de los tratamientos aplicados a los pacientes era más eficiente. Los resultados obtuvimos demostraron que el tratamiento con el medicamento Ursodiol proporciona una mayor supervivencia a las funciones del hígado.

Así mismo, se logró determinar las variables significativas para el análisis del mantenimiento del hígado. Donde, según el modelo de Cox, las variables más significativas son el tratamiento ofrecido a los pacientes, así como la puntuación de riesgo Mayo PBC que tiene el paciente. Demostrando que son 2 variables importantes para considerar en el análisis del fallo de hígado.

Por último, se logró determinar que los puntajes más altos registrados del riesgo de mayo en los pacientes poseen un menor índice de supervivencia según el análisis realizado. Demostrando así, que entre menor sea el puntaje de riesgo, mayor probabilidad de que el hígado no falle.

## Bibliografía

Dietmar, A. (2008). *Mayo risk score for primary biliary cirrhosis: a useful tool for the prediction of course after liver transplantation?* Mayo 15, 2021, de PubMed.gov Sitio web: <https://pubmed.ncbi.nlm.nih.gov/18806733/>

Nikolaos, T. (2017). *What determines the prognosis of primary biliary cholangitis (PBC)?* Mayo 15, 2021, de Medscape Sitio web: <https://www.medscape.com/answers/171117-79080/what-determines-the-prognosis-of-primary-biliary-cholangitis-pbc#:~:text=The%20Mayo%20risk%20score%20is,0%2C%200.5%2C%20or%201>

Anónimo. (2020). *Ursodiol*. Mayo 13, 2021, de InsideRx Sitio web: <https://insiderx.com/es/drugs/ursodiol>

Anónimo. (2017). *Cirrosis biliar primaria*. Mayo 13, 2021, de Medline Plus Sitio web: Anónimo. (2020). *Ursodiol*. Mayo 13, 2021, de InsideRx Sitio web: <https://insiderx.com/es/drugs/ursodiol>

Amat, J. (2016). *Regresión logística simple y múltiple*. Mayo 22, 2021, de RPubS Sitio web: [https://rpubs.com/Joaquin\\_AR/229736#:~:text=La%20Regresi%C3%B3n%20Log%C3%ADstica%20Simple%2C%20desarrollada,funci%C3%B3n%20de%20una%20variable%20cuantitativa.&text=La%20asignaci%C3%B3n%20final%20se%20hace%20en%20funci%C3%B3n%20de%20las%20probabilidades%20predichas](https://rpubs.com/Joaquin_AR/229736#:~:text=La%20Regresi%C3%B3n%20Log%C3%ADstica%20Simple%2C%20desarrollada,funci%C3%B3n%20de%20una%20variable%20cuantitativa.&text=La%20asignaci%C3%B3n%20final%20se%20hace%20en%20funci%C3%B3n%20de%20las%20probabilidades%20predichas).

Anónimo. (2016). *Conseguir este libro impreso▼ Portada 0 Opiniones Escribir un comentario Modeling Survival Data: Extending the Cox Model*. Mayo 21, 2021, de Google Books Sitio web: <https://books.google.com.mx/books?id=oj0mBQAAQBAJ&pg=PA180&lpg=PA180&dq=udca1+survival&source=bl&ots=rIJyUehjMv&sig=ACfU3U2Pi7c9gKkUo5M3iZUD8P1OPMG9uA&hl=es-419&sa=X&ved=2ahUKEwiL1eKmjtzwAhUMGKwKHSkSAiYQ6AEwBnoECBIAAw#v=onepage&q=udca1%20survival&f=false>



Anónimo. (2002). Qué es y cómo interpretar una regresión logística. Mayo 22, 2021, de Conceptos claros Sitio web: <https://conceptosclaros.com/que-es-regresion-logistica/>