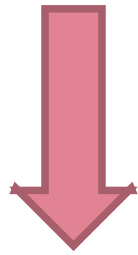


# Análisis de supervivencia

**Variable principal**



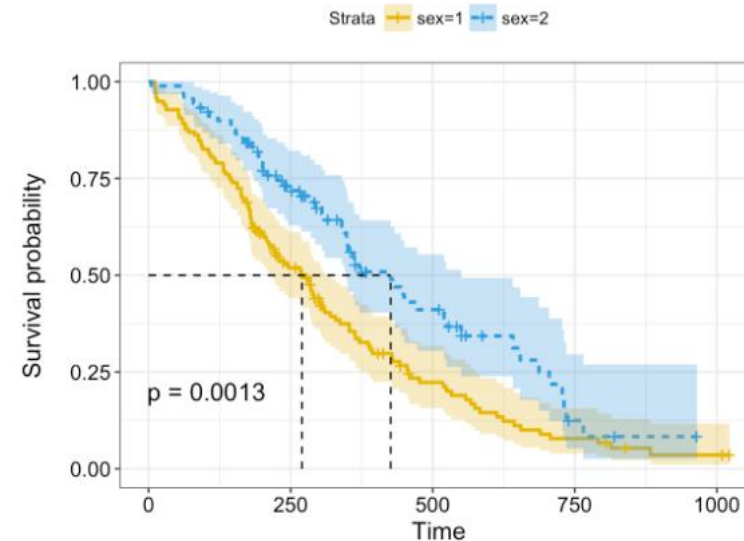
tiempo que tarda en ocurrir un suceso, ya sea este **beneficioso** (alta hospitalaria) o **perjudicial** (muerte, recaída), o **indiferente** (cambio de tratamiento).



**Análisis de Supervivencia**

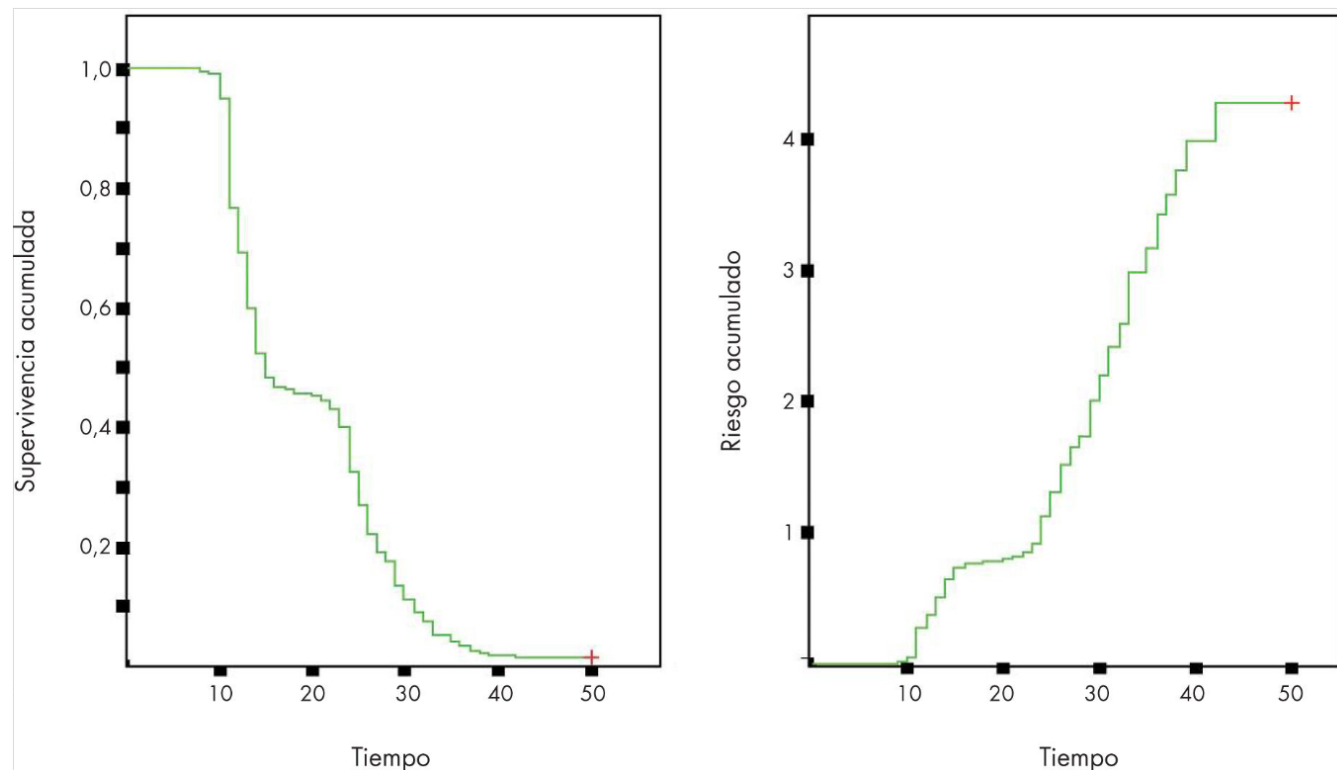
**Cuando usarlo...**

muy apropiada para analizar respuestas binarias (sí/no; fallecido/vivo) en estudios longitudinales



# Supervivencia y riesgo

- La supervivencia es la probabilidad de que un individuo sobreviva (no ocurra el evento) desde la fecha inicio de seguimiento hasta un momento determinado en el tiempo  $t$ .
- El riesgo es la probabilidad de que un individuo que está siendo observado en el tiempo  $t$  muera (suceda el evento) en ese momento.



# Puntos clave

Datos faltantes:

- **Censuras:** Pérdidas de seguimiento o fin del estudio.
- **Truncamientos:** Entrada en el estudio después del hecho que define el origen.

En lugar de ignorar completamente a estos individuos, se los considera como parte del conjunto de datos, pero su contribución se limita al período de tiempo durante el cual se tuvo información sobre su supervivencia.

Metodología estadística:

Paramétricas	Semiparamétricas	No Paramétricas
M.Exponencial	Regresión de Cox	Kaplan-Meier S[T]
Modelo Weibull	M.R.P. de Aalen	Curvas de Nelson-Aalen
M.Log-Normal M.Log-Logístico	M.R.C. de Fine y Gray	Log-Rank S[T]

**Métodos no paramétricos:** Describir la forma general de la función de supervivencia sin hacer supuestos específicos sobre su forma.

**Paramétricos:** Para comparar modelos específicos de distribución de supervivencia o si deseas hacer predicciones a largo plazo.

# Estudio

Los datos sobre el día de la muerte de 863 pacientes con trasplante de riñón. A todos los pacientes se les realizó su trasplante.

El tiempo máximo de seguimiento para este estudio fue de 9.47 años. Los pacientes fueron censurados si se mudaron o si estaban vivos al final del seguimiento.

En la muestra, había 432 hombres blancos, 92 hombres negros, 280 mujeres blancas y 59 mujeres negras.

Las edades de los pacientes en el trasplante variaron de 9,5 meses a 74,5 años con una edad promedio de 42,8 años.

Setenta y tres (16.9%) de los hombres blancos, 14 (15.2%) de los hombres negros, 39 (13.9%) de las mujeres blancas y 14 (23.7%) de las mujeres negras murieron antes del final del estudio.

1. **obs**. Numero de observación.
2. **time**. Tiempo hasta la muerte o en estudio.
3. **delta**. Indicador de muerte(1=Vivo,0=Muerte)
4. **gender**. 1=Hombre, 2=Mujer
5. **race**. Raza (1=Blanco,2=Negro)
6. **age**. Edad en Años.

	obs	time	delta	gender	race	age
1	1	1	0	1	1	46
2	2	5	0	1	1	51
3	3	7	1	1	1	55
4	4	9	0	1	1	57
5	5	13	0	1	1	45
6	6	13	0	1	1	43
7	7	17	1	1	1	47
8	8	20	0	1	1	65
9	9	26	1	1	1	55
10	10	26	1	1	1	44
Showing 1 to 11 of 863 entries, 6 total columns						

# Exploración

Número de observaciones censuradas de acuerdo a las variables



```
> with(kidtran, table(delta)) #Censura
delta
 0    1
723 140
```



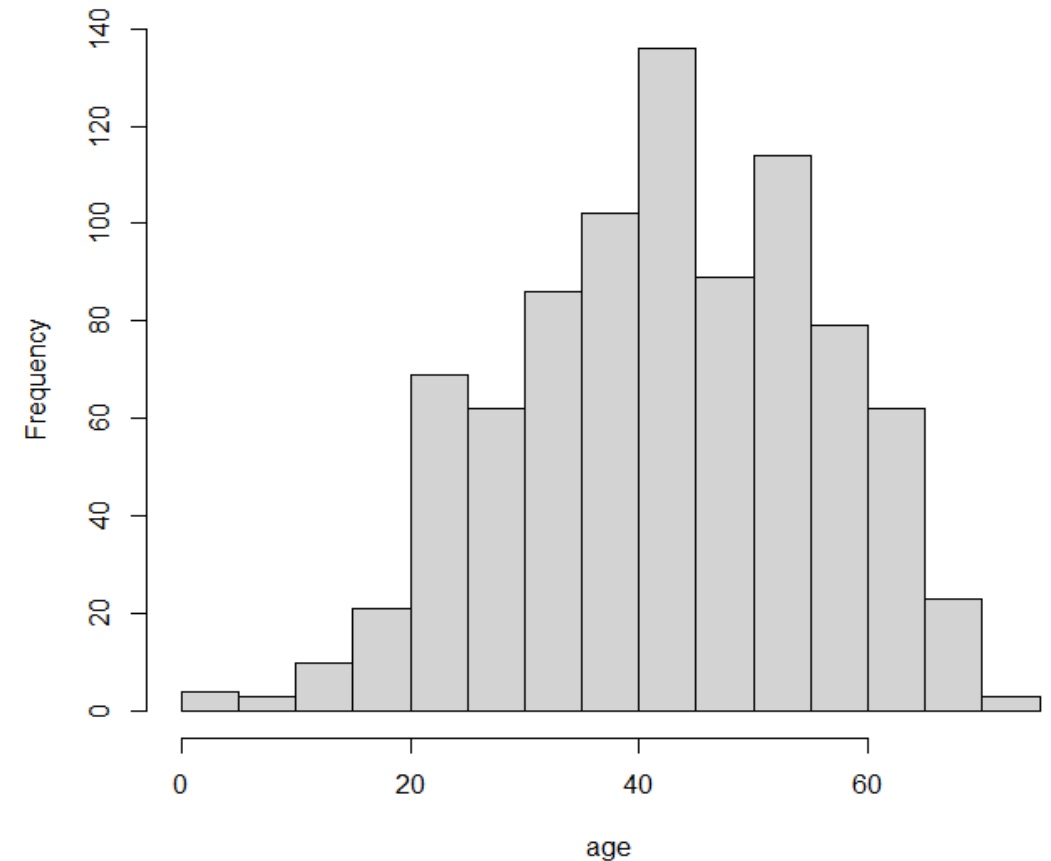
```
> with(kidtran, prop.table(table(delta)))
delta
      0      1
0.8377752 0.1622248
```



	delta	
	0	1
gender		
Hombre	437	87
Mujer	286	53

	delta	
	0	1
race		
Blanco	600	112
Negro	123	28

Histogram of age

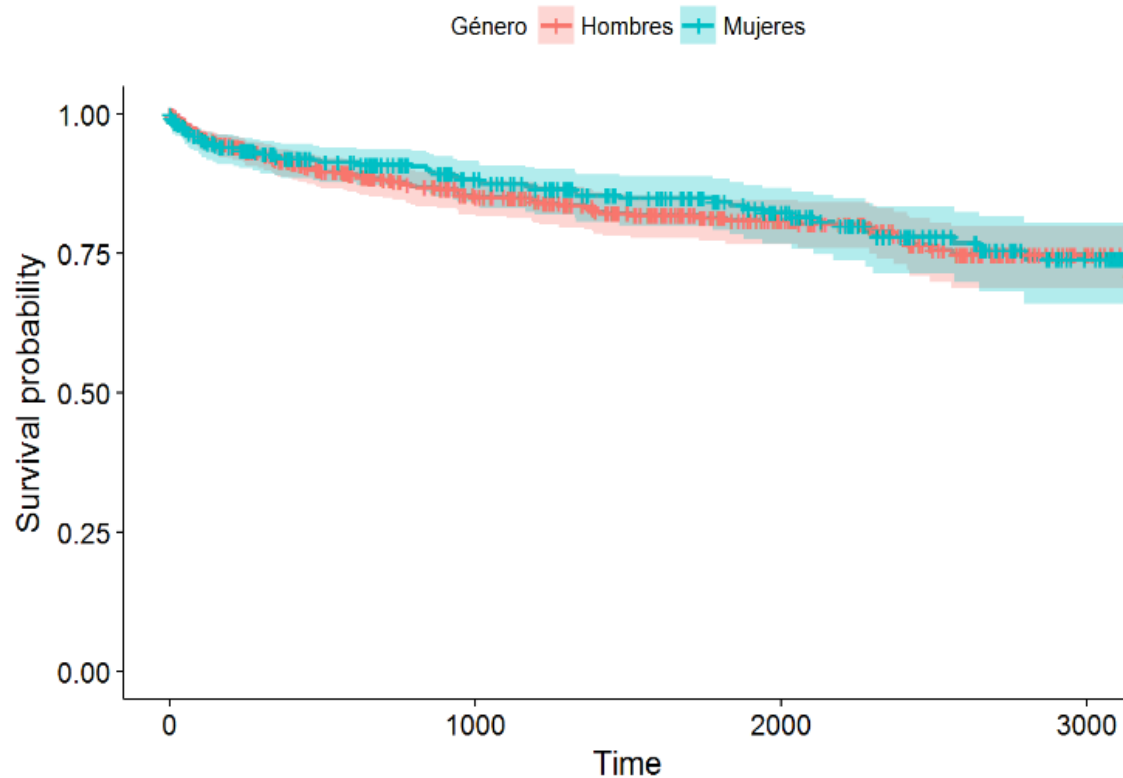


```
> kidtran$edadg <- cut(kidtran$age, c(0, 25, 50, 75))
> with(kidtran, table(edadg, delta))
      edadg      delta
      0      1
(0,25]    100     7
(25,50]   406    69
(50,75]   217    64
```

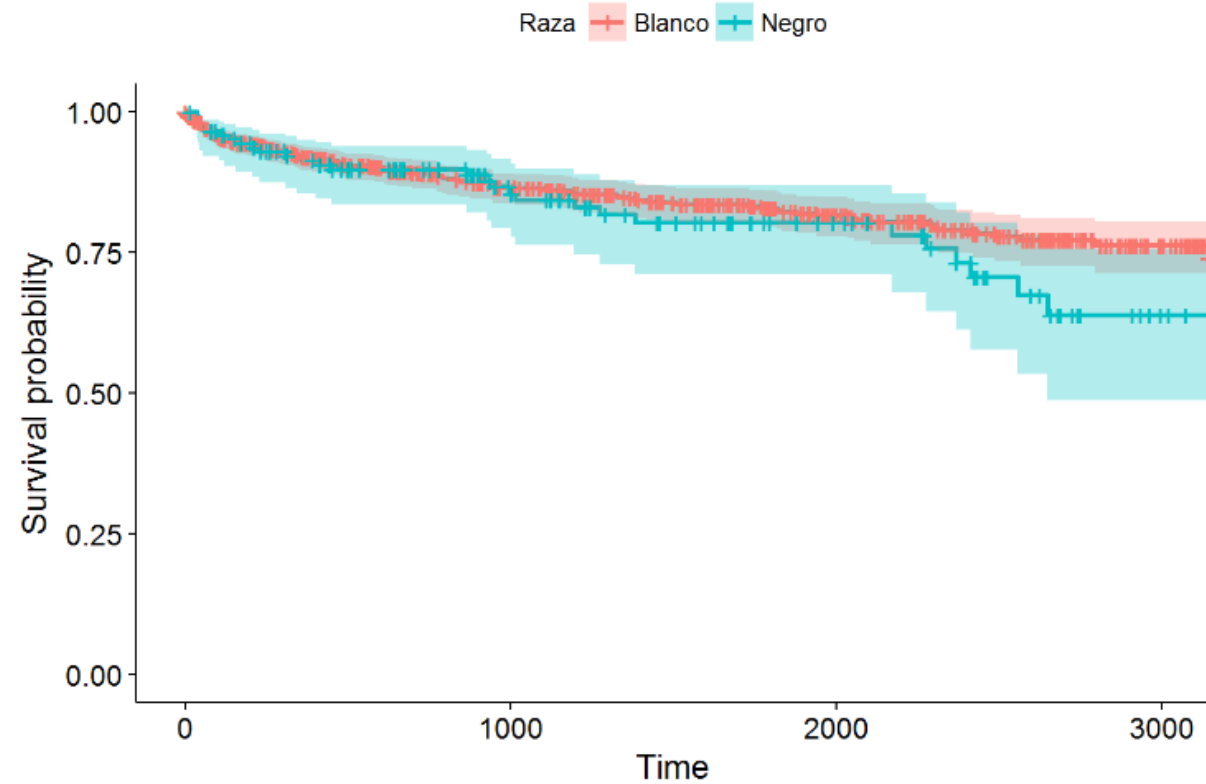
# 1.- Estimación No-Paramétrica

Se estimará la curva de supervivencia para diferentes variables con la finalidad de encontrar alguna diferencia notable entre estas. (fleming-harrington o KM)

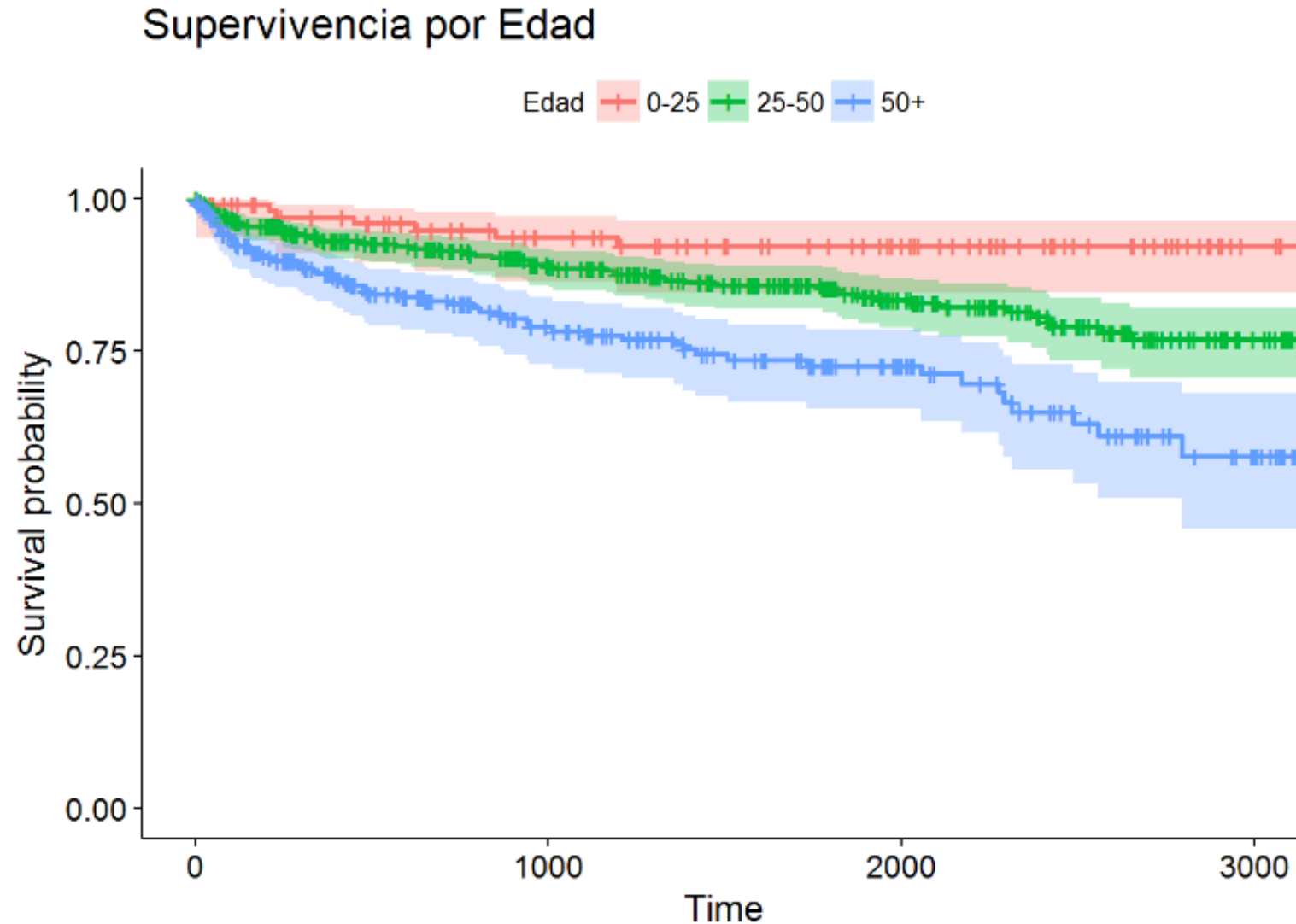
Supervivencia por Género



Supervivencia por Raza

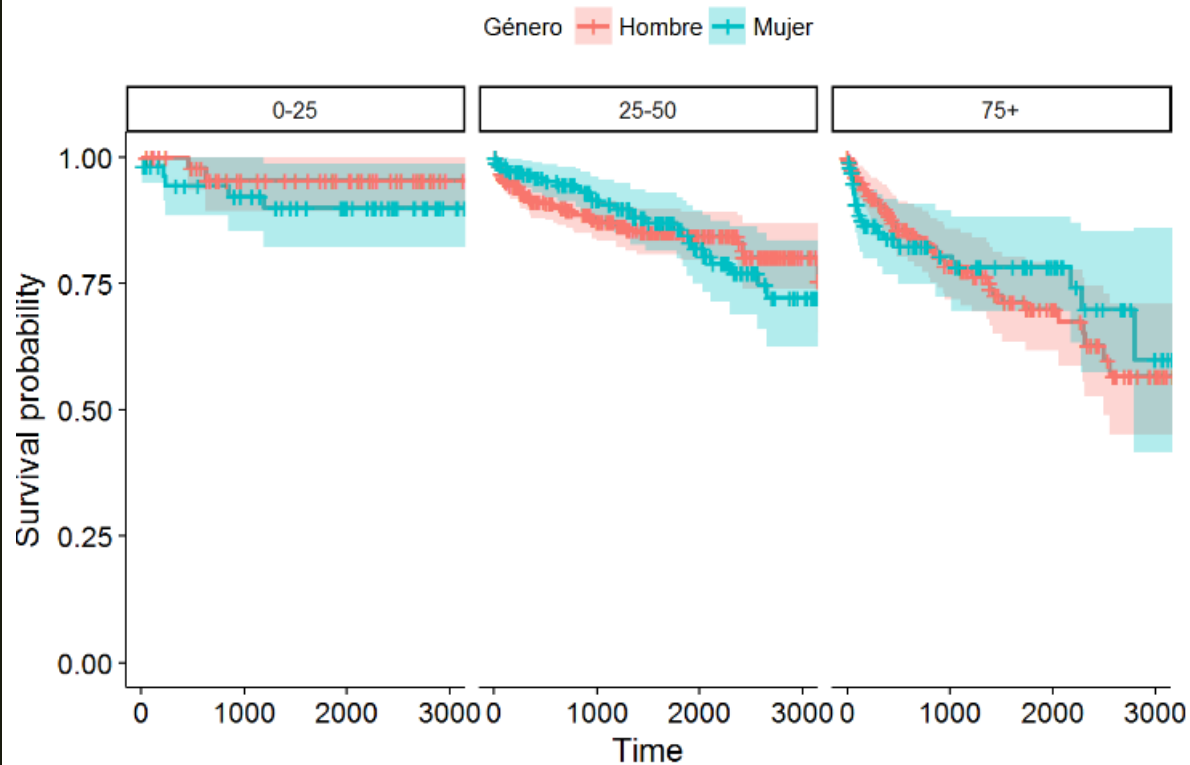


A primera vista los grupos de edad presentan una clara diferencia entre sus curvas de supervivencia, en donde las personas con mayor edad presentan un mayor riesgo, la curva por Raza presenta al final una ligera diferencia.

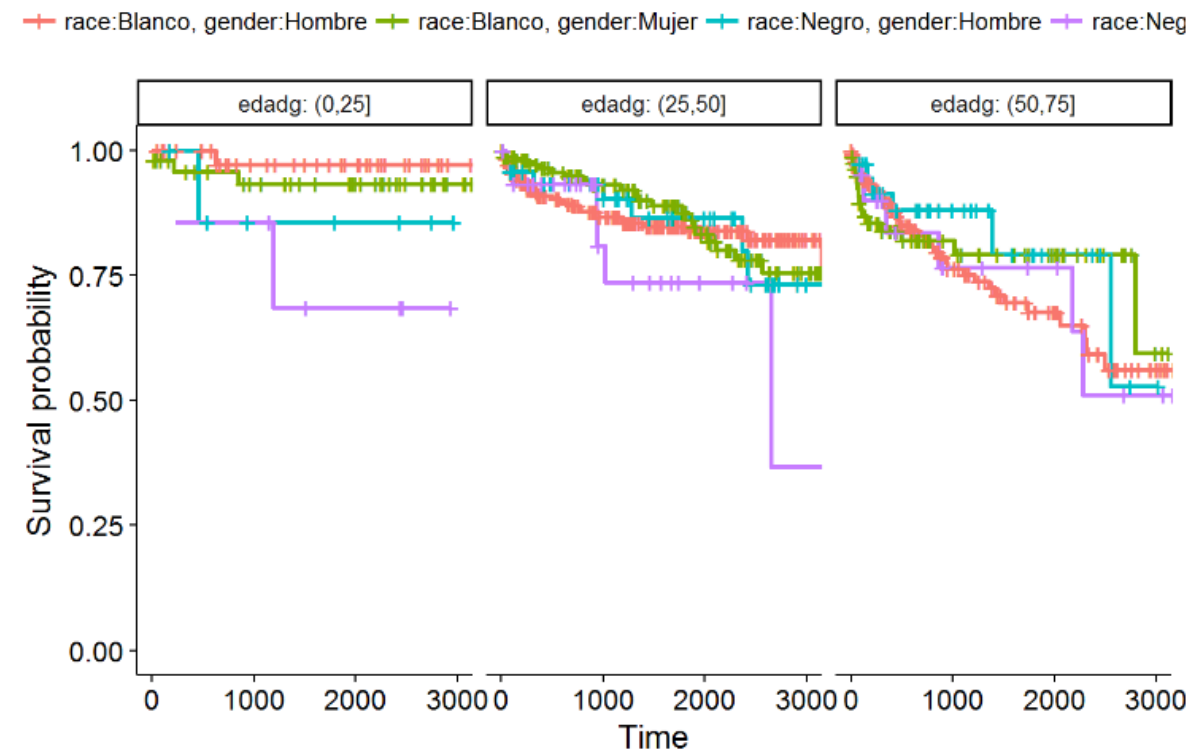


# Con más de 1 variable (covariable)

Supervivencia entre Género por Edad



Supervivencia entre Género por Raza



La edad gráficamente parecer un factor de riesgo, esta hipótesis se pondrá a prueba a continuación.



# Pruebas de Hipótesis

1

Pruebas de hipótesis de funciones de supervivencia (log-Rank o Peto & Peto):

1. Género y Raza
2. Género y Raza estratificando los grupos de Edad.
3. Grupos de Edad.
4. Tendencia entre los Grupos de Edad.

```
## Call:
## survdiff(formula = Surv(time, delta) ~ gender + race, data = kidtran,
##          rho = 1)
##
##
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
## gender=Hombre, race=Blanco	432	66.3	62.71	0.2114	0.462
## gender=Hombre, race=Negro	92	12.5	13.16	0.0342	0.042
## gender=Mujer, race=Blanco	280	35.2	42.62	1.2777	2.131
## gender=Mujer, race=Negro	59	12.5	8.04	2.4179	2.827

```
##
## Chisq= 4.3 on 3 degrees of freedom, p= 0.227
```

2

```
## Call:
## survdiff(formula = Surv(time, delta) ~ gender + race + strata(edadg),
##          data = kidtran, rho = 1)
##
##
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
## gender=Hombre, race=Blanco	432	64.9	63.91	0.0149	0.034
## gender=Hombre, race=Negro	92	12.3	14.01	0.1972	0.248
## gender=Mujer, race=Blanco	280	35.3	38.74	0.3009	0.490
## gender=Mujer, race=Negro	59	12.4	8.25	2.0389	2.451

```
##
## Chisq= 2.9 on 3 degrees of freedom, p= 0.413
```

3

```
## Call:
## survdiff(formula = Surv(time, delta) ~ edadg, data = kidtran,
##          rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## edadg=(0,25] 107      6.43     18.7      8.08     10.60
## edadg=(25,50] 475     61.87     72.8      1.63      4.22
## edadg=(50,75] 281     58.24     35.0     15.34     23.40
##
##  Chisq= 27.8  on 2 degrees of freedom, p= 9.19e-07
```

En 1 y 2 podemos suponer que las variables género y raza no son factores de riesgo para la muerte en el trasplante de riñón.

No rechazan la hipótesis de diferencia

Se **rechaza** la hipótesis de igualdad de curvas de supervivencia entre los grupos de edad 3.

Log-rank puede indicar si hay diferencias entre las curvas de supervivencia, pero no proporciona información sobre qué grupos específicos difieren ni sobre la magnitud del efecto.

Si la prueba de log-rank resulta significativa, se pueden realizar análisis adicionales, como la regresión de Cox

## 2.- Estimación Paramétrica

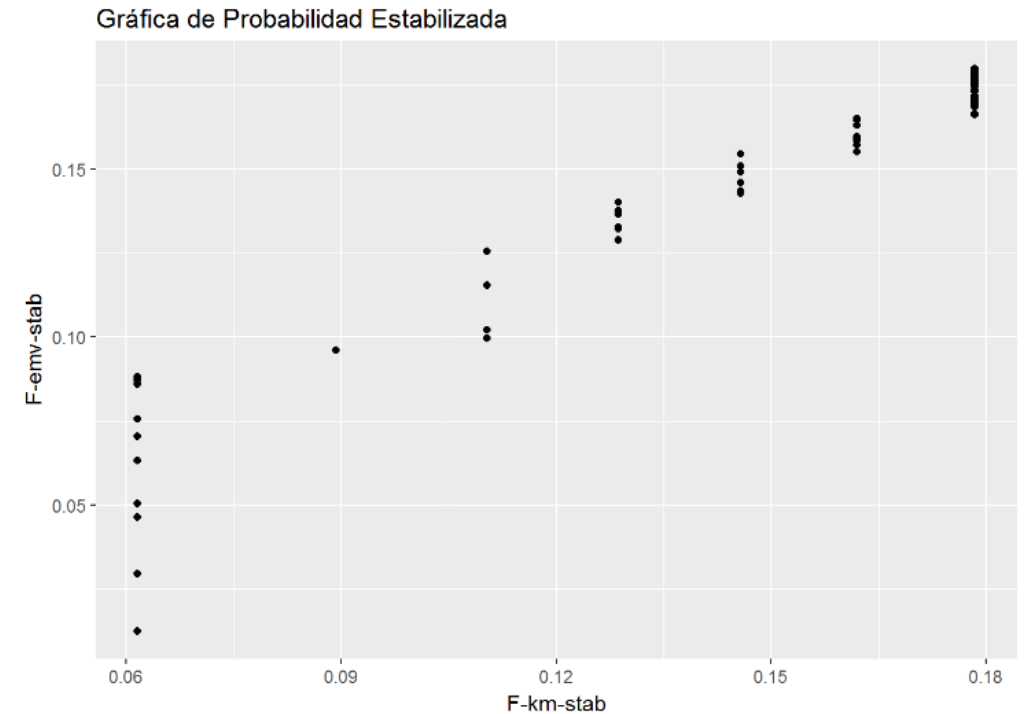
Se dividirá el dataset en los 3 grupos de edad.

Para cada grupo se estimará las siguientes distribuciones: Weibull, Loglogística, Gompertz y LogNormal.

Eligiéremos al modelo que presente un mejor AIC y compararemos con las curvas de supervivencia observadas.

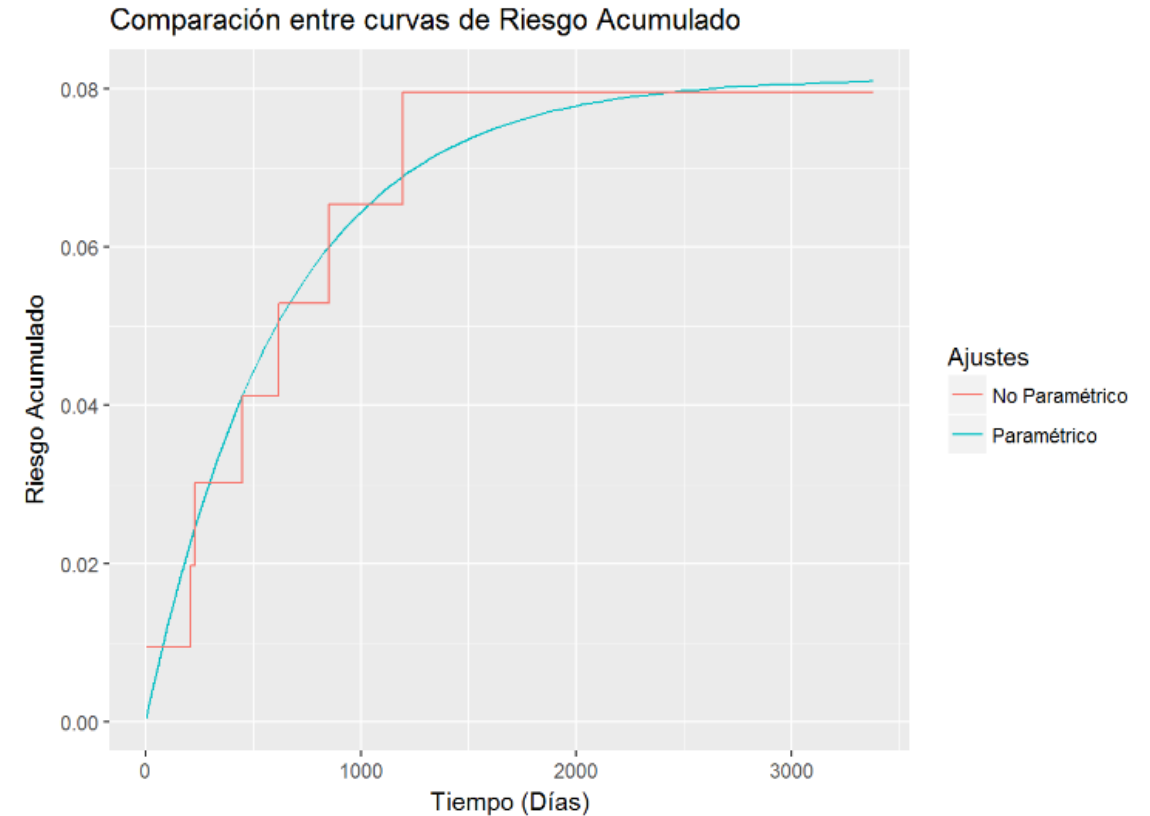
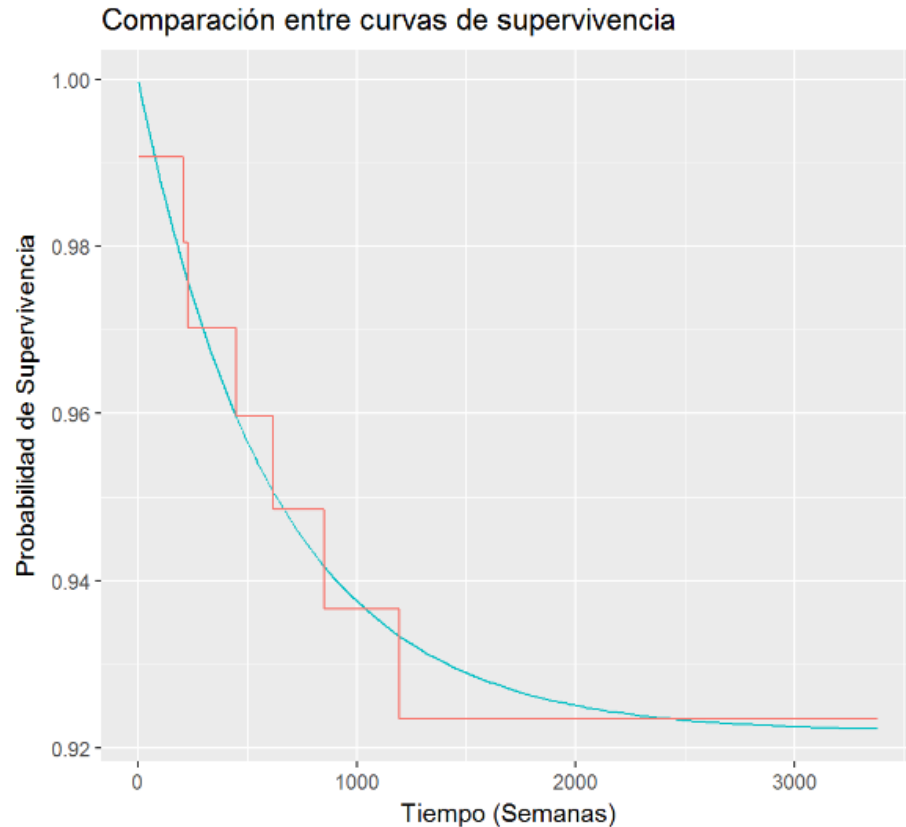
```
##          gompertz lnorm llogis weibull
## AIC          155    155    155     156
## BIC          160    161    161     161
## LogLik       -75    -76    -76     -76
```

Ahora **validaremos** con el Gráfico de Probabilidad Estabilizada:



Tendencia de línea recta, pero con un gran porcentaje de observaciones censuradas se presentan líneas verticales.

## Visualización de curva de supervivencia y riesgo acumulado.



Al graficar la función de supervivencia y la función de riesgo estimadas a partir del modelo paramétrico y compararlas con las estimaciones no paramétricas, se puede evaluar visualmente si el modelo paramétrico captura bien la forma de la distribución de supervivencia

➡ Inferencia

# Modelo Semiparamétrico

Selección de Variables (Regresión de Cox):

```
## Call:
## coxph(formula = Surv(time, delta) ~ age + race + gender, data = kidtran)
##
##               coef exp(coef) se(coef)      z      p
## age           0.05104   1.05236  0.00718  7.11 1.2e-12
## raceNegro     0.11636   1.12341  0.21151  0.55  0.58
## genderMujer  0.02654   1.02689  0.17490  0.15  0.88
##
## Likelihood ratio test=57.1  on 3 df, p=2.5e-12
## n= 863, number of events= 140
```