

# **Defunciones de menores de una semana por causas**

**Luis Miguel Calvo Magaz y Adrián Poncela Gómez**

**24 de Abril de 2019**

# **Datos elegidos: Defunciones de menores de una semana por causas (lista perinatal), sexo y edad.**

Los datos que hemos utilizado para la realización de este trabajo de Análisis de supervivencia, se trata de un conjunto de datos obtenidos del INE. Apartado: Sociedad-Salud-Defunciones según la Causa de Muerte 2017

Link: [http://www.ine.es/jaxi/Tabla.htm?  
path=/t15/p417/a2017/l0/&file=01006.px&L=0](http://www.ine.es/jaxi/Tabla.htm?path=/t15/p417/a2017/l0/&file=01006.px&L=0)

Esos datos muestran, el número de fallecidos por cada día, dentro de la primera semana de vida de los bebés. Hay que tener en cuenta, que todos los individuos que se estudian, fallecen en este estudio. No se estudia ningún individuo que viva más de una semana.

Los datos se muestran, divididos por sexo y edad de fallecimiento.

La edad de fallecimiento está calculada por diferencia entre la fecha de defunción y la de nacimiento, a excepción de los menores de 24 horas de vida. Por tanto, como fallecidos de un día figuran sólo aquellos que han vivido más de 24 horas y fallecido al día siguiente.

Variable objeto de estudio: Edad (días) de fallecimiento de los bebés que no vivieron más de una semana

Origen: 1 de Enero de 2017

Suceso de interés: Fallecimiento del bebé.

Escala del tiempo: 1 de Enero de 2017 - 31 de Diciembre de 2017

Censuras y truncamientos: Para este conjunto de datos, no tenemos censuras ni truncamientos. Todos los individuos que son objeto de estudio, mueren, ninguno sobrevive, ya que todos los nacidos que viven más de una semana no son parte de este estudio.

Dentro de este conjunto de datos, además del sexo, se nos proporciona otra covariable, que es la causa de muerte.

Las causas de muerte aparecen divididas en apartados que son los que hemos utilizado, ya que considerar más de 90 casos de muerte para tan pocos datos (490 en un año), no sería relevante.

Los apartados de causas de muerte considerados, siguiendo la división que utiliza el INE, son los siguientes:

1. Perinatal (383 individuos)
2. Malformaciones (86 individuos)
3. Enfermedades recién nacido (20 individuos)
4. Causas externas (1 individuo)

Sin embargo, sólo un bebé se considera dentro del apartado de “Causas externas”. Visto que no hay una similitud, de este individuo con las demás causas, y al ser sólo una observación, hemos decidido prescindir de él para la realización del estudio.

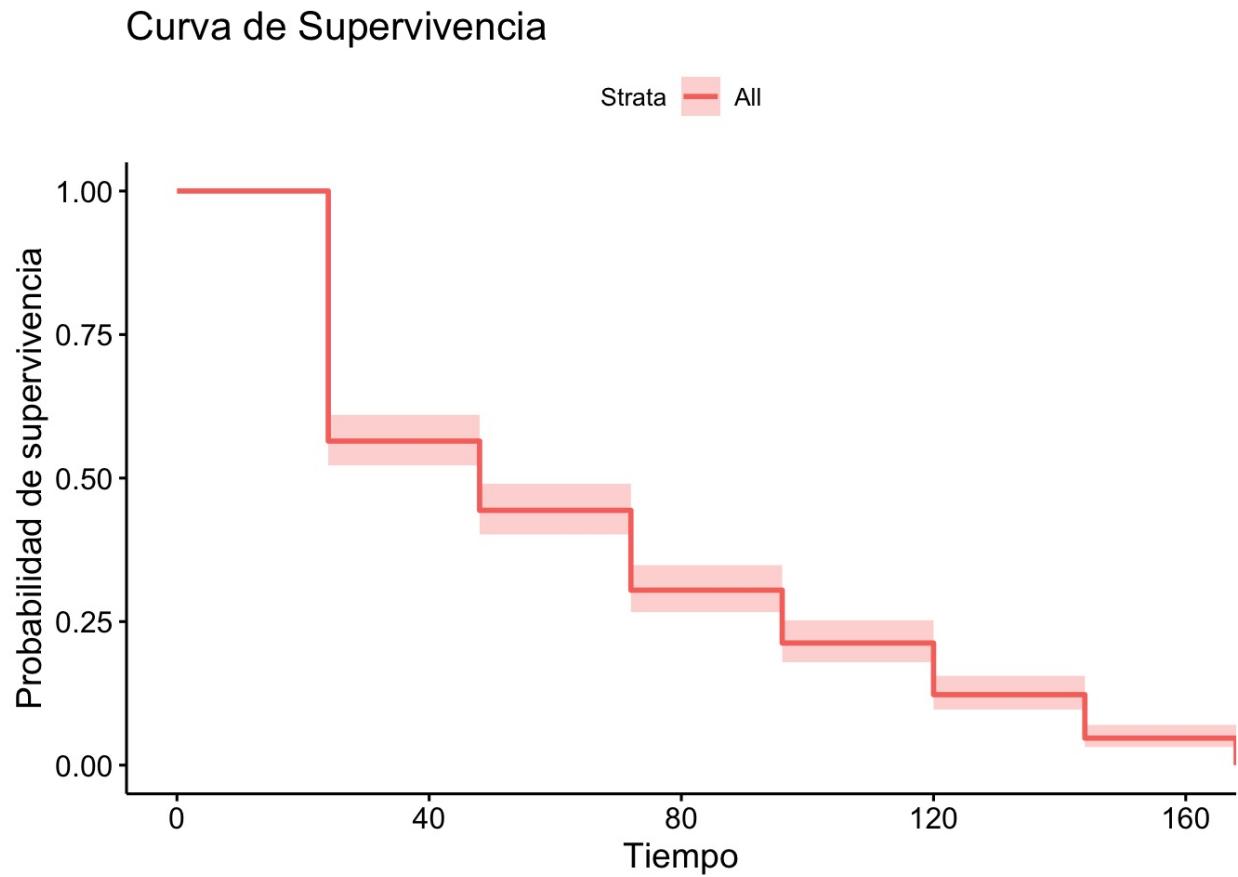
Respecto a los 3 tipos de causa que nos quedan, vemos como la distribución de los individuos en cada una de las causas, es es igual, ya que una causas tiene más individuos que las demás. Se ve que hay una gran diferencia entre ellas. Hay 4 veces más individuos muertos por malformaciones que por enfermedades de recien nacido. Además Hay casi 5 veces más muertos por perinatal que por malformaciones.

Para realizar el estudio,partiendo de los datos del INE, se han generado 4 ficheros de datos que son los siguientes:

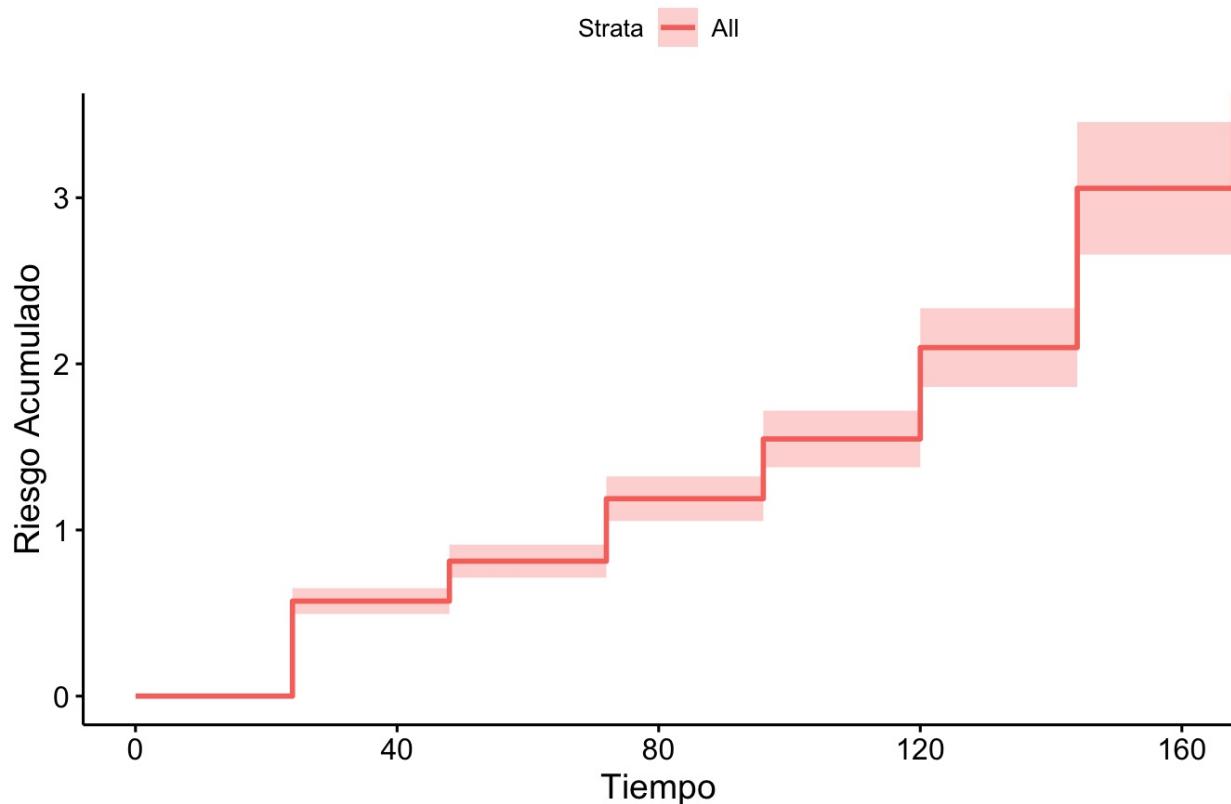
1. todas\_las\_causas.csv Datos por sexo y edad con todas las causas de muerte.
2. fichero\_datos\_1.csv Datos por sexo y edad con causa de muerte: Perinatal
3. fichero\_datos\_2.csv Datos por sexo y edad con causa de muerte: Malformaciones
4. fichero\_datos\_3.csv Datos por sexo y edad con causa de muerte: Enfermedades recién nacido

# Estimación Kaplan Meier sin covariables

```
## Call: survfit(formula = mortalidad.superv ~ 1, data = datos_mortalidad)
##          type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##      n      events      *rmean    *se(rmean)      median      0.95LCL
##      489.00     489.00      64.69       2.08      48.00      48.00
##      0.95UCL
##      48.00
##      * restricted mean with upper limit = 168
```



## Riesgo Acumulado



Podemos ver cómo en las primeras 24 horas se alcanza el mayor número de defunciones de recién nacidos. A partir de ahí, el número de muertes disminuye, decreciendo linealmente.

Los intervalos de confianza se estrechan a medida que avanza el tiempo. El número de muertes, por cada día, disminuye con el paso del tiempo. La curva finaliza con valor cero, ya que todos los recién nacidos de este estudio, han fallecido.

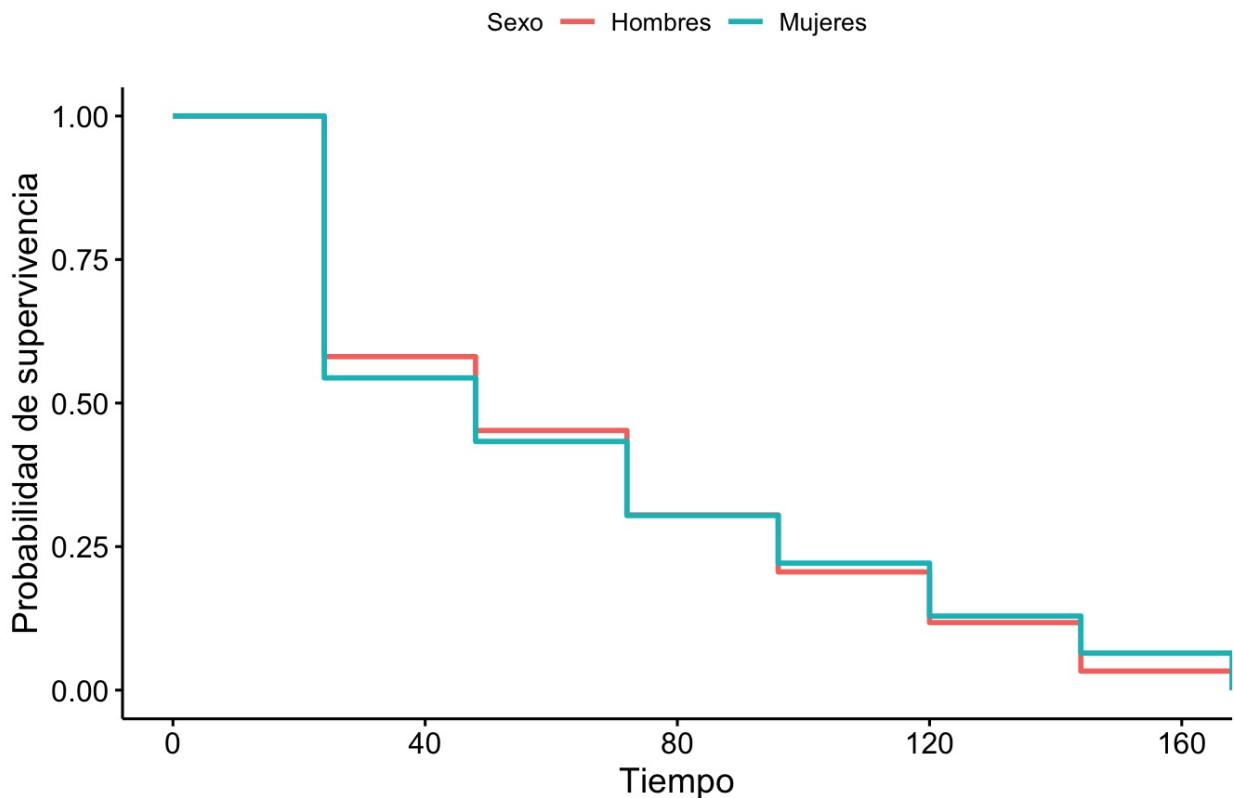
En cuanto al grafico de la tasa de riesgo, se ve que los intervalos de confianza a medida que avanza el tiempo, aumentan su amplitud. Destacando, que en día 7 se muestra un Intervalo bastante amplio (Duda).

El primer incremento de riesgo es grande, luego baja considerablemente, y a partir de ahí sube hasta alcanzar el mayor incremento de riesgo al séptimo día.

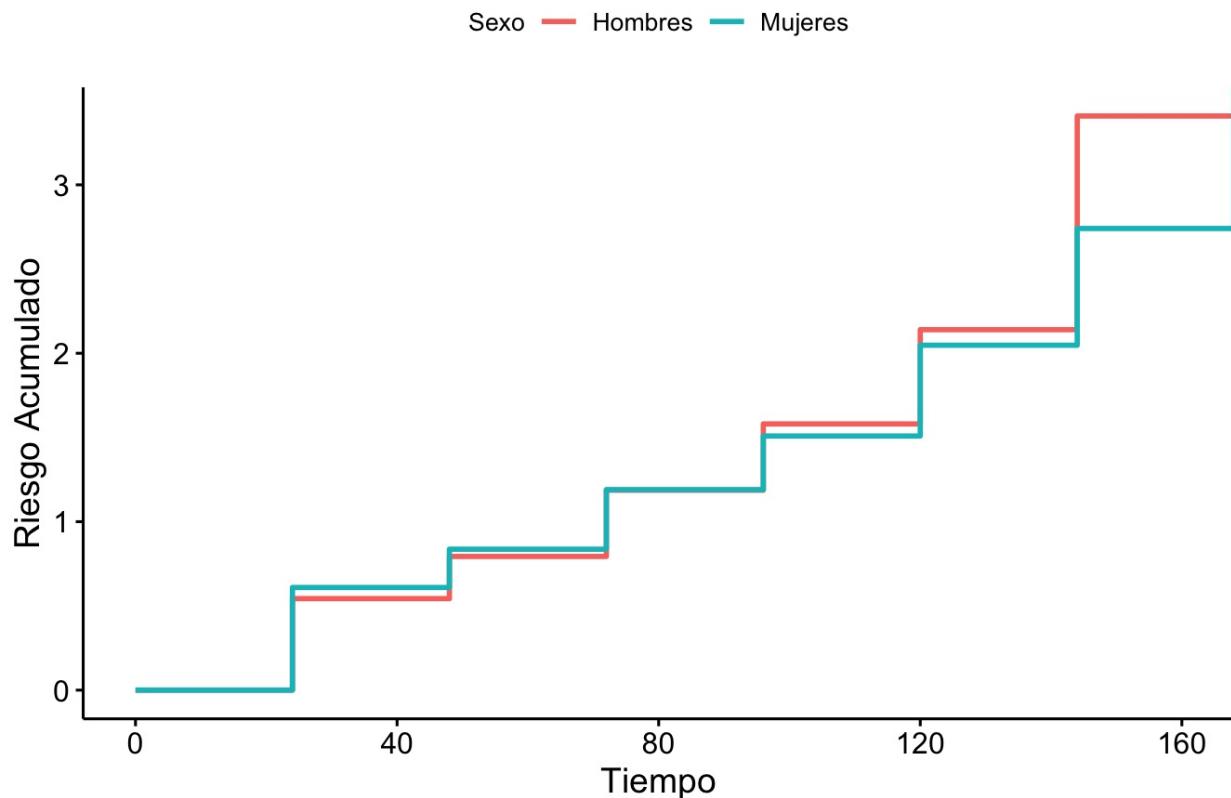
# Estimación Kaplan Meier según el sexo

```
## Call: survfit(formula = mortalidad.superv ~ Sexo, data = datos_mc)
##           type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##          n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## Sexo=    Hombres 272     272   64.7      2.70     48      48
## Sexo=    Mujeres 217     217   64.7      3.22     48      24
##          * restricted mean with upper limit = 168
```

Curva de Supervivencia



## Riesgo Acumulado



Curva similar a la ya explicada, sin Covariables. Decrecimiento grande al principio, y a partir de ahí, disminuye linealmente.

La diferencia entre sexos, no es apreciable, ya que las dos gráficas son muy parecidas.

Destacar que en el segundo y tercer día la probabilidad de sobrevivir era mayor en los hombres que en las mujeres.

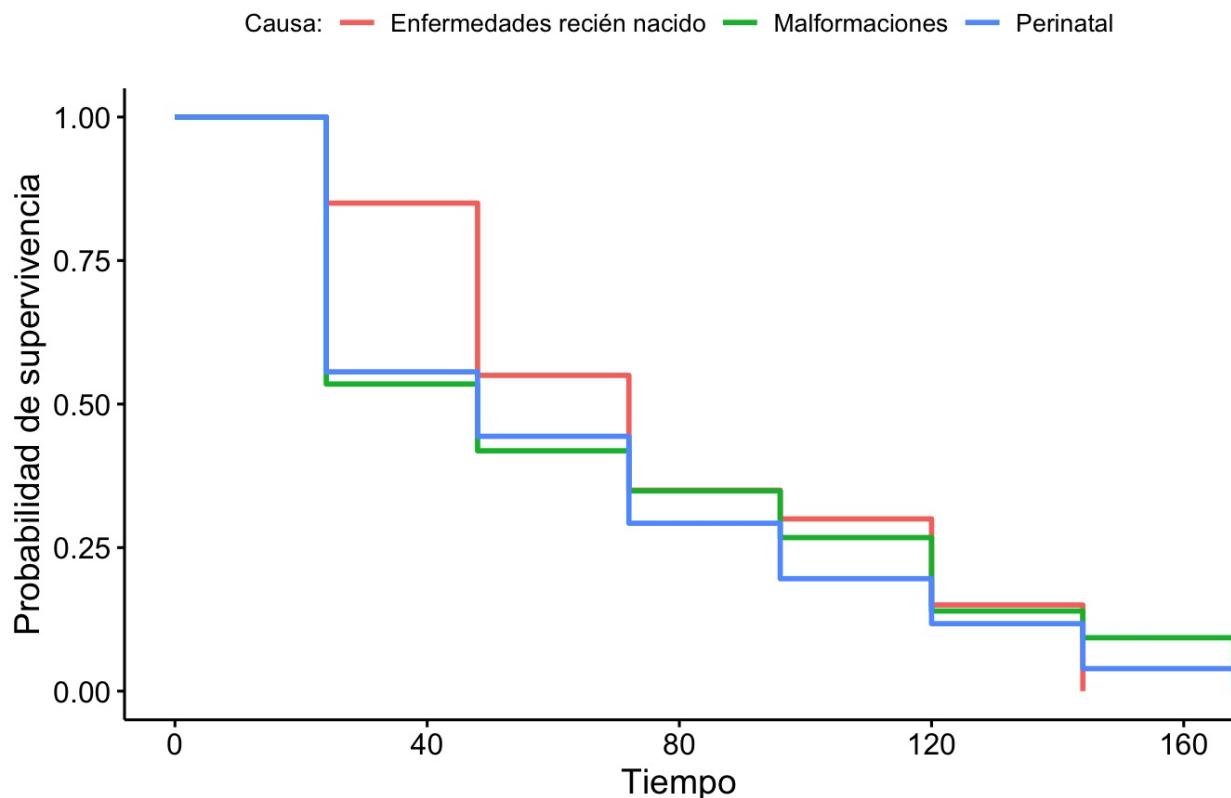
El cuarto día, la probabilidad es la misma, y los últimos 3 días se invierte; es decir la probabilidad de sobrevivir era mayor en las mujeres que en los hombres. Incremento lineal, muy similar entre ambos sexos.

Apenas hay diferencias gráficamente, quitando las del séptimo.

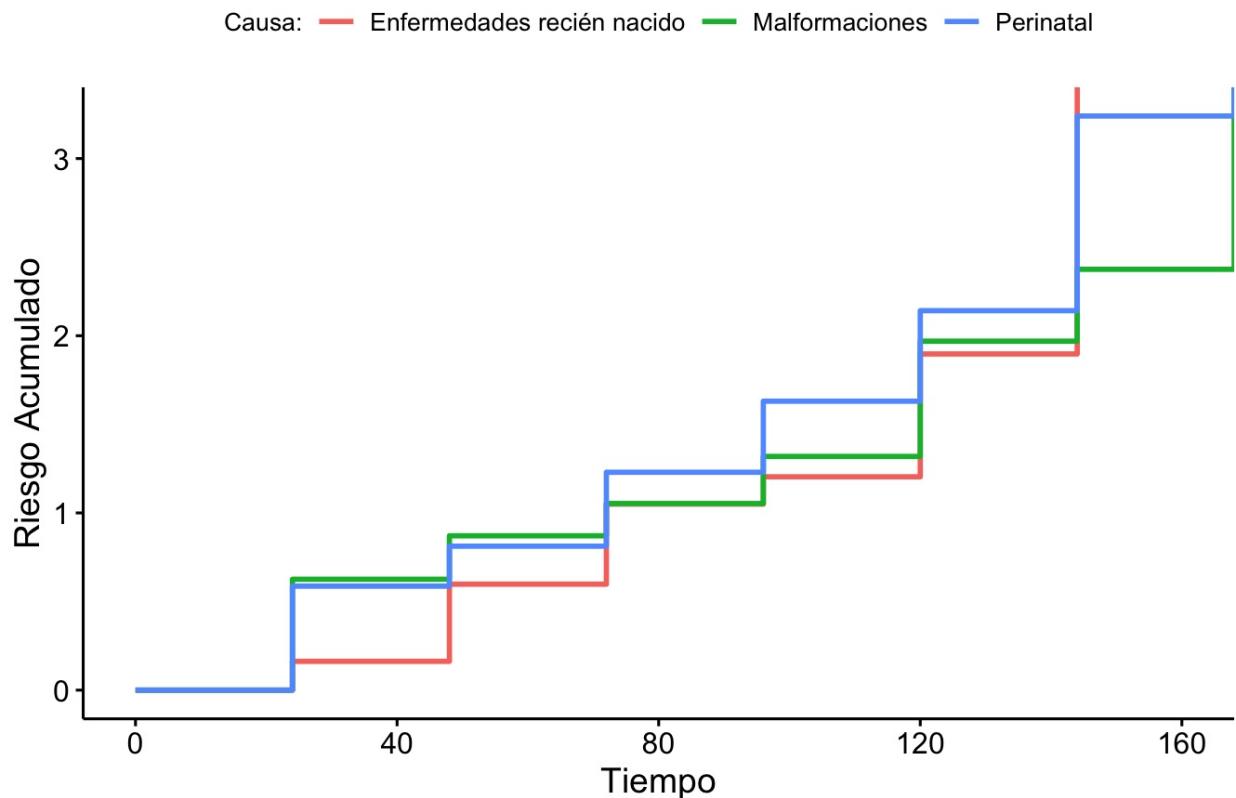
# Estimación Kaplan Meier según la causa

```
## Call: survfit(formula = mortalidad.superv ~ causa, data = datos_m
##                 type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##                                         n events *rmean *se(rmean) med
## causa=Enfermedades recién nacido    20      20   76.8      9.08
## causa=Malformaciones                86      86   67.3      5.45
## causa=Perinatal                   383     383   63.5      2.30
##                                         0.95LCL 0.95UCL
## causa=Enfermedades recién nacido    48      120
## causa=Malformaciones                24      72
## causa=Perinatal                   48      48
##      * restricted mean with upper limit = 168
```

Curva de Supervivencia



## Riesgo Acumulado



El individuo que murió por causas externas es atípico, no tiene mucha relevancia en el estudio, lo suprimimos del estudio. En este gráfico, podemos ver dos partes. Una que es antes de las 72 horas, que serían los 3 primeros días. Y de los 3 días en adelante, es decir, más de 72 horas. En la primera parte del gráfico se ve como hay dos curvas muy similares, que son las de Malformaciones y Perinatal, que están muy juntas. Sin embargo, la tercera causa que se corresponde con las enfermedades del recién nacido, va por encima, más adelante veremos si es significativamente diferente o no.

En la segunda parte del gráfico, se ve cómo la curva de enfermedades de recién nacido, ya no se separa de las otras dos, sino que las tres curvas, son bastante similares.

# Distribuyendo aleatoriamente los valores de tiempo uniformes (0,24), (24,48), ...

Decidimos aumentar la complejidad, del estudio.

Los datos les teníamos agrupados según la muerte por día. Hemos hecho una simulación, utilizando la distribución uniforme. La idea es distribuir las muertes, a lo largo del día, como si fuesen la edad de muerte, ya no sólo por el día de la muerte, sino por la hora de la muerte, distribuyendo el número de muertes por cada día, a lo largo de las 24 horas de ese día.

## Distribución uniforme de defunciones

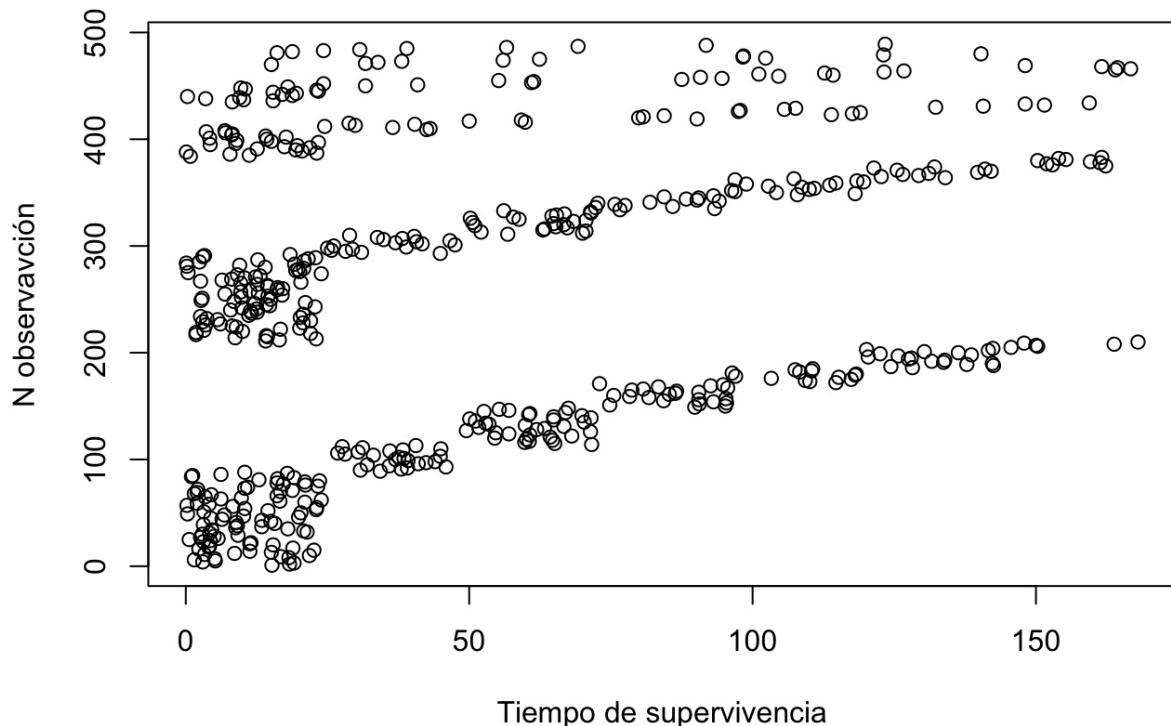
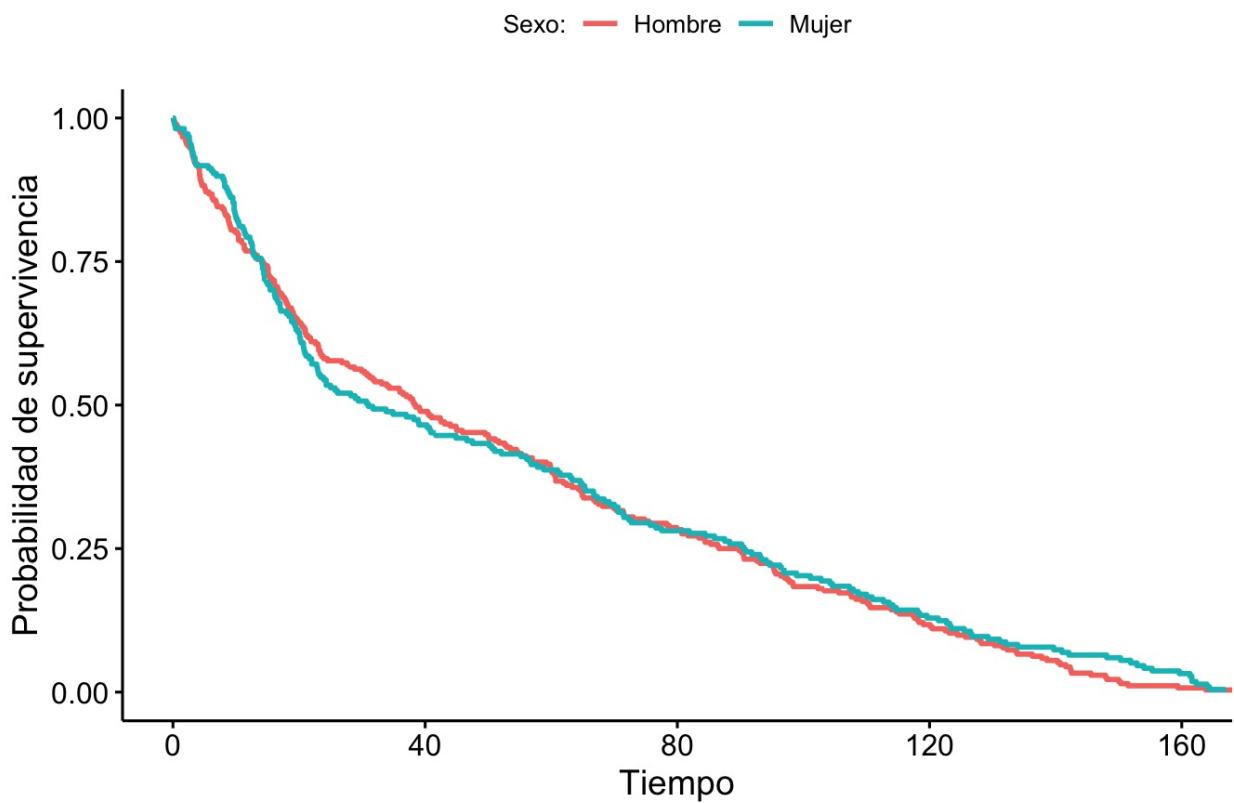


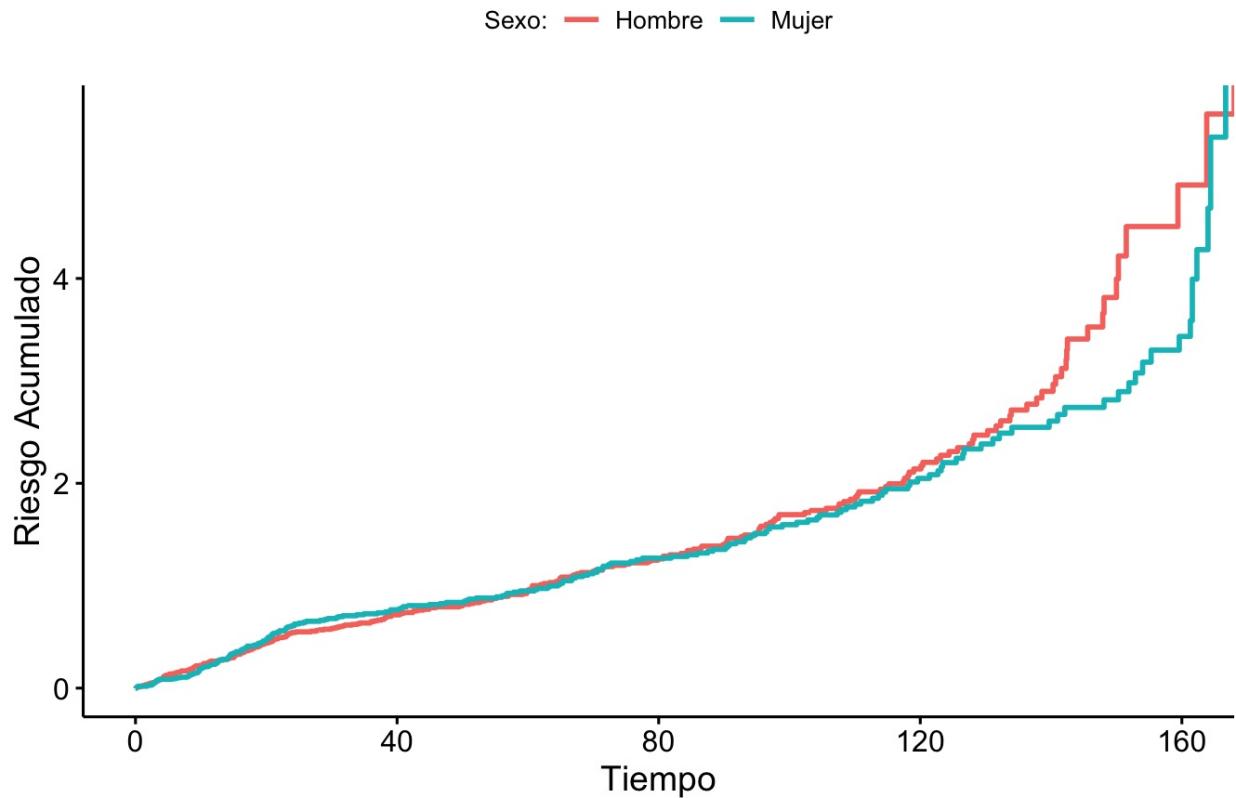
Gráfico de la distribución de los datos, utilizando la uniforme.

```
## Call: survfit(formula = mortalidad.superv_rdm_sexo ~ Sexo, data =
##                 type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##          n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## Sexo=    Hombres 272     272   52.7       2.76   38.4    30.4    5
## Sexo=    Mujeres 217     217   53.2       3.26   31.0    22.9    5
##          * restricted mean with upper limit =  167
```

## Curva de Supervivencia



## Riesgo Acumulado



Los resultados, con la simulación de la uniforme, son diferentes en cuanto a las gráficas anteriores, a que ahora ya no son escalonadas, ya no muestran saltos, por cada día. Se muestra una función más lineal, sin saltos, aunque manteniendo las interpretaciones que concluimos anteriormente.

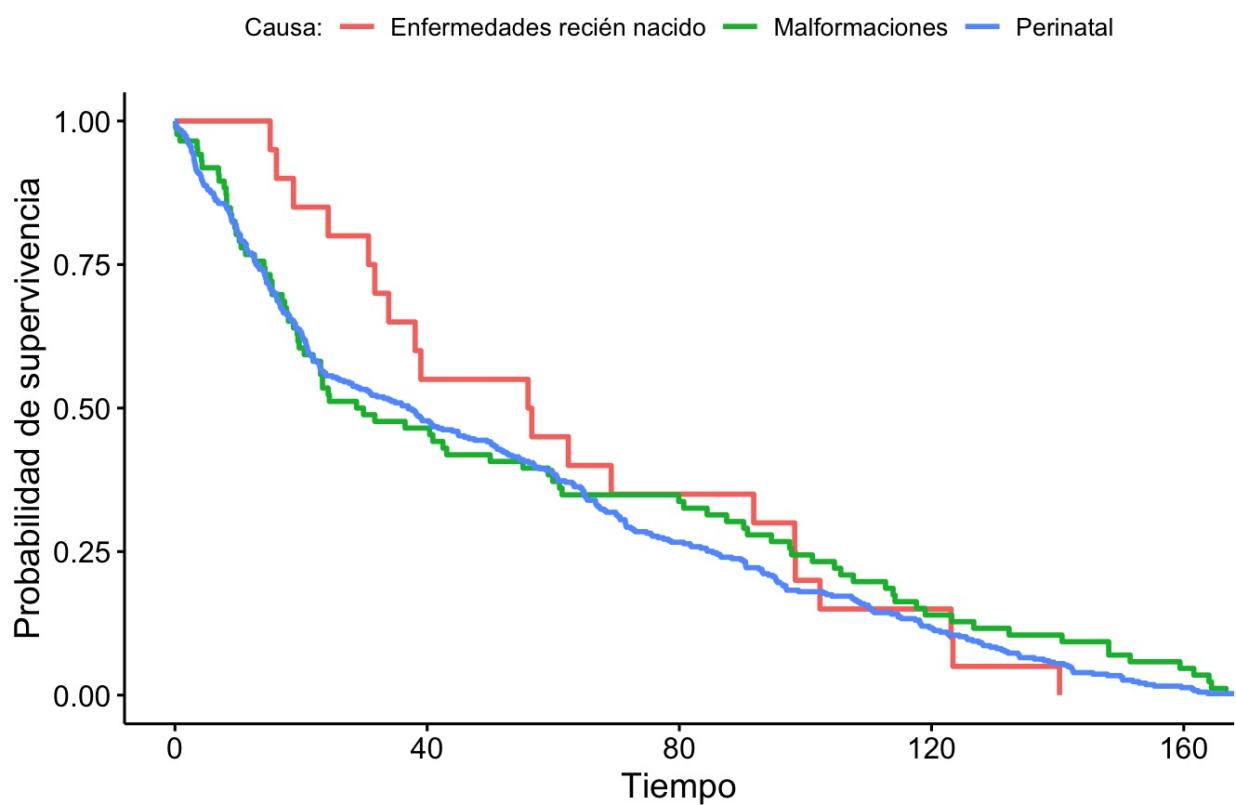
Apenas hay diferencias, entre hombres y mujeres. Ésto es un dato que llama la atención. Ya que, podría existir una diferencia más considerable en cuanto al número de fallecidos en la priemra semana.

Vemos que no es así.

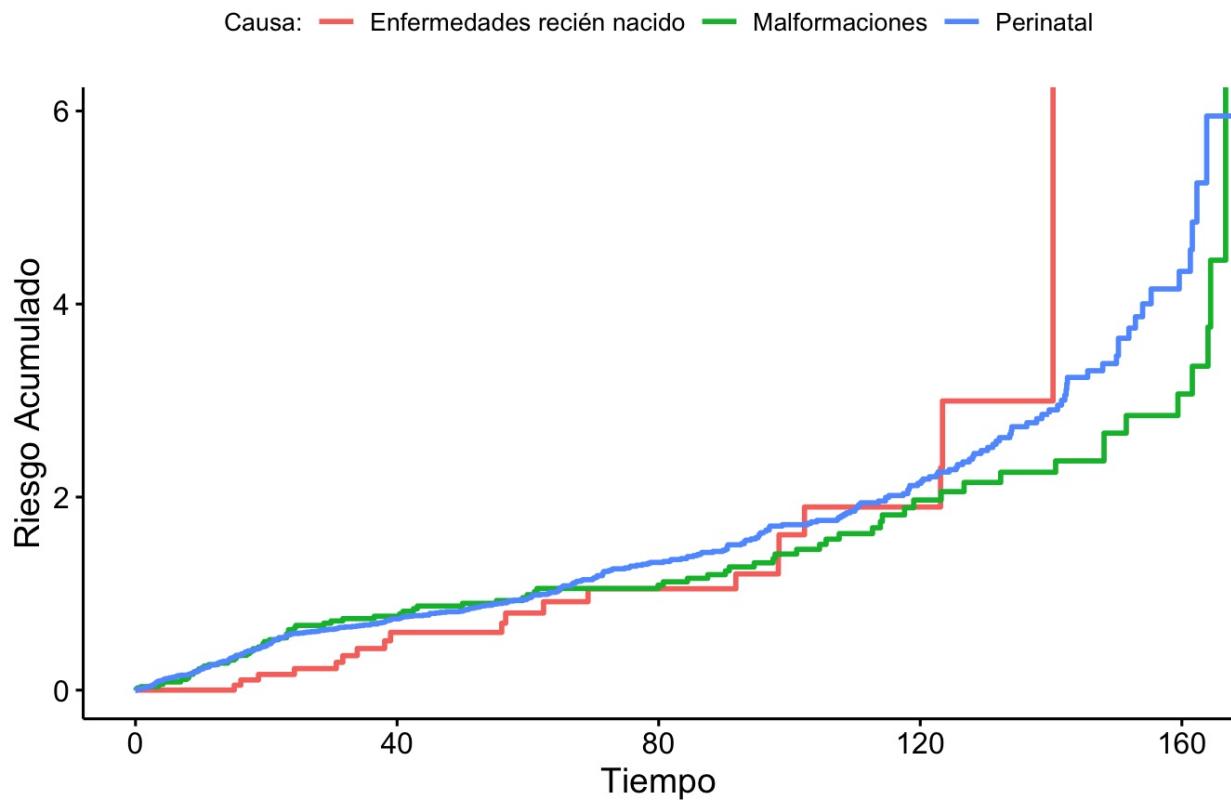
```
## Call: survfit(formula = mortalidad.superv_rdm ~ causa, data = dat
##                 type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##                                         n events *rmean *se(rmean) med
## causa=Enfermedades recién nacido    20      20   63.5      8.70  5
## causa=Malformaciones                86      86   55.6      5.53  2
## causa=Perinatal                     383     383   51.8      2.34  3
##                                         0.95LCL 0.95UCL
## causa=Enfermedades recién nacido    33.9     98.4
```

```
## causa=Malformaciones      20.5    59.9
## causa=Perinatal          26.8    46.6
##   * restricted mean with upper limit = 167
```

## Curva de Supervivencia



## Riesgo Acumulado



En este gráfico, vemos lo que ya comentamos anteriormente. El gráfico se puede dividir en dos. Aproximadamente en antes y después de 72 horas. En la primera parte, dos curvas muy similares que son las de perinatal y malformaciones; ambas diferentes de la de enfermedades del recién nacido. A partir de las 72 horas, las 3 curvas se asemejan más, no apreciando grandes diferencias visualmente.

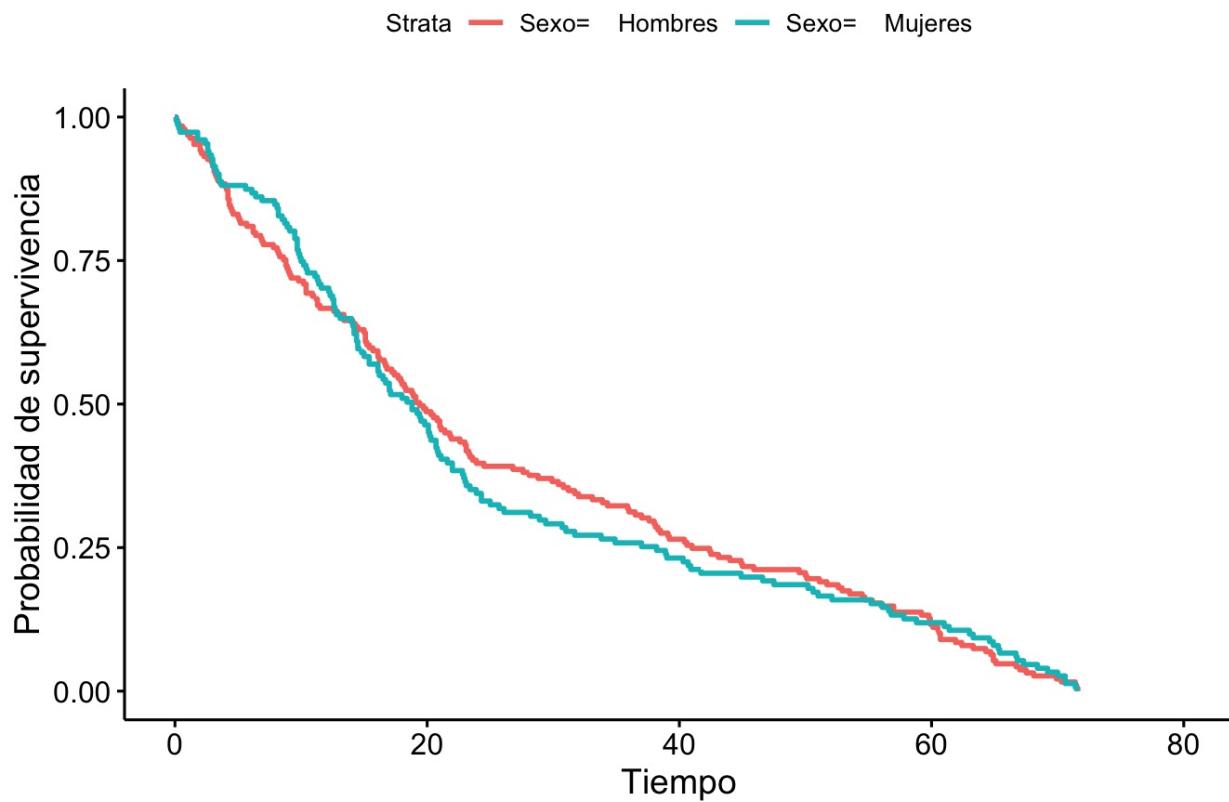
## Nos centramos en las muertes antes de las 72 horas

### Sexo

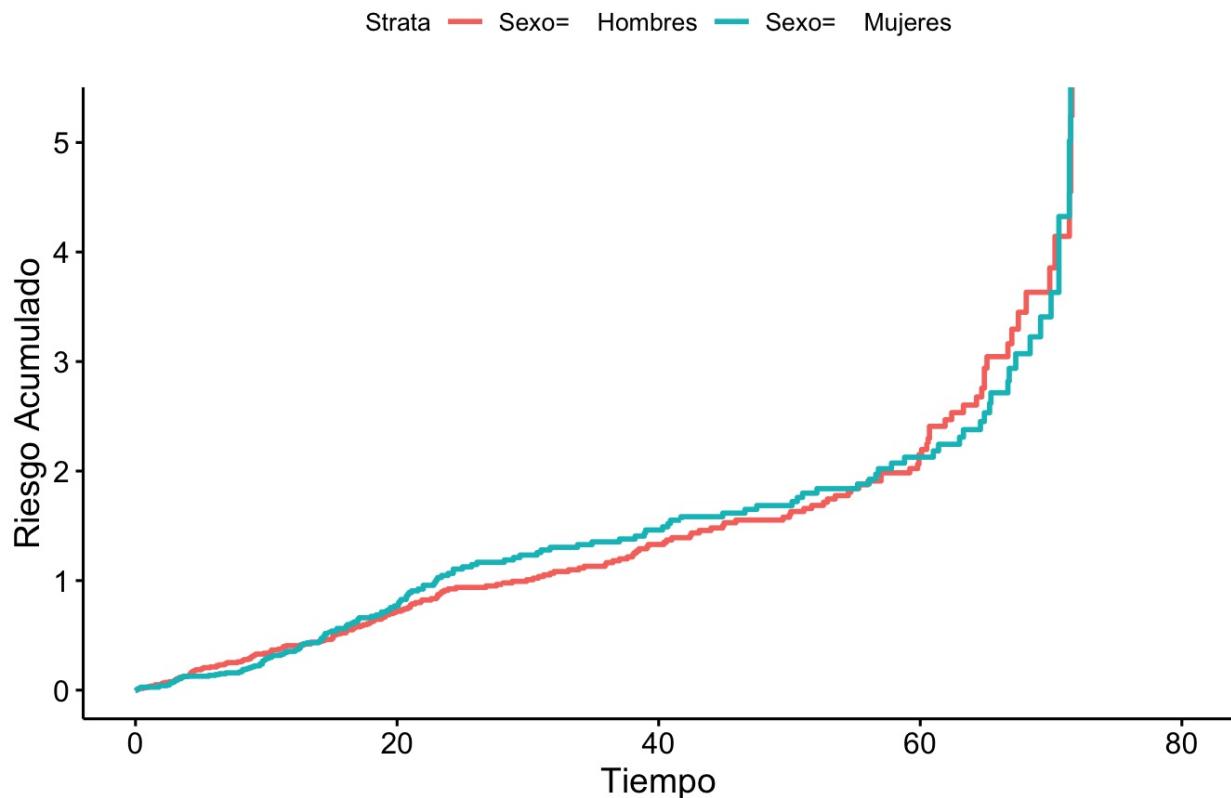
```
## Call: survfit(formula = mortalidad.superv_rdm_72 ~ Sexo, data = d
##                 type = "kaplan-meier")
##
##      1 observation deleted due to missingness
##          n events   *rmean   *se(rmean) median 0.95LCL 0.95UCL
##  Sexo=    Hombres 189     189    26.3      1.54    19.4    16.7    2
##  Sexo=    Mujeres 151     151    25.3      1.67    18.8    15.4    2
```

```
##      * restricted mean with upper limit =  71.5
```

## Curva de Supervivencia



## Riesgo Acumulado



Contraste log rank para los grupos de Sexo (Hombres y mujeres)

```
survdiff(mortalidad.superv_rdm_72~Sexo, data = datos_mortalidad_t_ale

## Call:
## survdiff(formula = mortalidad.superv_rdm_72 ~ Sexo, data = datos_
##
## n=340, 1 observation deleted due to missingness.
##
##          N Observed Expected (0-E)^2/E (0-E)^2/V
## Sexo=    Hombres 189      189     191   0.0144   0.0334
## Sexo=    Mujeres 151      151     149   0.0184   0.0334
##
##  Chisq= 0 on 1 degrees of freedom, p= 0.9
```

Causa

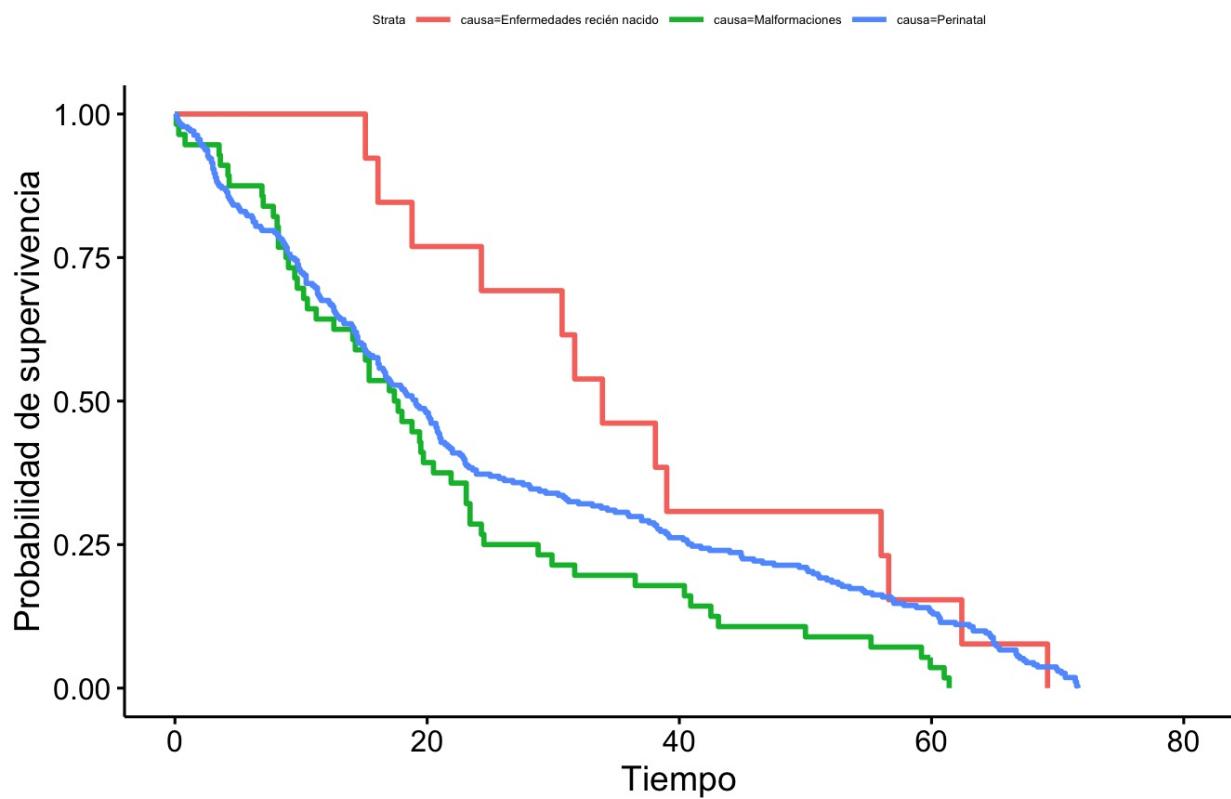
```
## Call: survfit(formula = mortalidad.superv_rdm_72 ~ causa, data =
##               type = "kaplan-meier")
##
## 1 observation deleted due to missingness
```

```

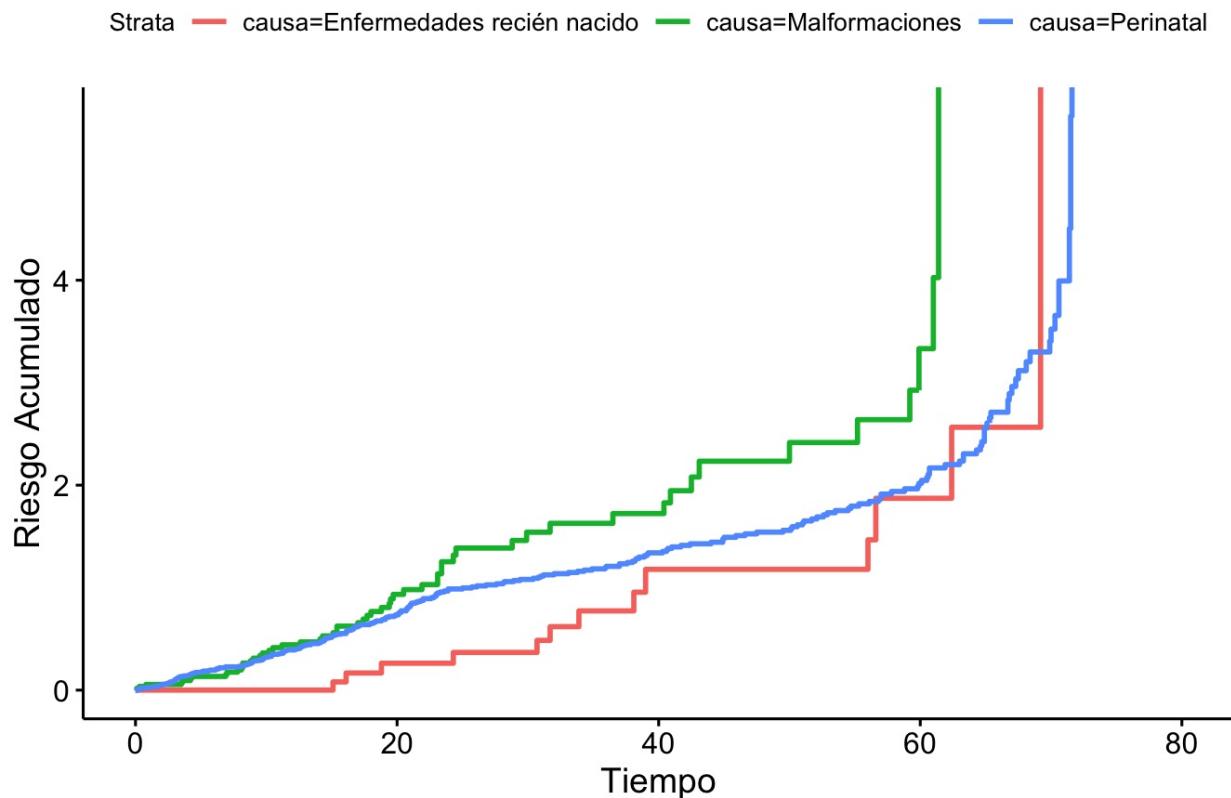
##                                     n  events *rmean *se(rmean) med
## causa=Enfermedades recién nacido 13      13  37.8      4.81   3
## causa=Malformaciones           56      56  21.3      2.20   1
## causa=Perinatal                271     271  26.2      1.30   1
##                                         0.95LCL 0.95UCL
## causa=Enfermedades recién nacido 24.3      NA
## causa=Malformaciones           14.1     23.1
## causa=Perinatal                 16.5     21.0
##      * restricted mean with upper limit =  69.2

```

## Curva de Supervivencia



## Riesgo Acumulado



Contraste log rank para las covariables de causa sin tener indiferentemente del tiempo.

```
survdiff(mortalidad.superv_rdm~causa, data = datos_mortalidad_t_aleat

## Call:
## survdiff(formula = mortalidad.superv_rdm ~ causa, data = datos_mc
##
## n=489, 1 observation deleted due to missingness.
##
##                               N Observed Expected (O-E)^2/E
## causa=Enfermedades recién nacido 20      20    22.5     0.283
## causa=Malformaciones            86      86    96.6     1.154
## causa=Perinatal                383     383   369.9     0.463
##
##  Chisq= 1.9  on 2 degrees of freedom, p= 0.4
```

Contraste log rank para las covariables de causa para los individuos que fallecen antes de 72 horas.

```
survdiff(mortalidad.superv_rdm_72~causa, data = datos_mortalidad_t_al
```

```

## Call:
## survdiff(formula = mortalidad.superv_rdm_72 ~ causa, data = datos
##
## n=340, 1 observation deleted due to missingness.
##
##                                     N Observed Expected (0-E)^2/E
## causa=Enfermedades recién nacido 13        13     18.5    1.651
## causa=Malformaciones            56        56     42.4    4.397
## causa=Perinatal                271       271    279.1    0.236
##
##  Chisq= 6.4  on 2 degrees of freedom, p= 0.04

```

## Estimador de COX

El modelo de COX plantea el logaritmo del riesgo relativo como una función lineal de las variables independientes. Se supone, por lo tanto, que el riesgo relativo, a diferencia del riesgo propiamente dicho, no depende del tiempo o, dicho de otra manera, que es constante a lo largo del tiempo (de ahí el nombre de modelo de riesgo proporcional).

Esto no se cumple en nuestro conjunto de datos ya que, como podemos ver, el riesgo aumenta exponencialmente, por lo tanto, el modelo de COX no es el modelo adecuado para nuestro conjunto.