

# Riesgos\_Competitivos

## Introducción y preeliminares en el análisis de datos de un hospital: Impacto del estado del estado de neumonía al momento de la admisión en la unidad de terapia intensiva

En este trabajo, abordamos el análisis de supervivencia de riesgos competitivos, por lo que el esquema de trabajo cambia un poco respecto al visto en clase, que es realizar el modelo de Cox. Así mismo, como tenemos otro enfoque (Al tratar ahora desde el punto de vista de riesgos competitivos), no aplican las estimaciones de Kaplan-Meier, sino otros estimadores.

Trabajaremos con la base sir.adm; disponible en el paquete “mvna”, llamado así por las iniciales de “Multivariate Nelson-Aalen estimator”, que permite estimar de manera no paramétrica los riesgos acumulados de transición de modelos de Markov multi-estados arbitrarios, usando dicho estimador.

```
library(mvna) #Estimadores de Nelson-Aalen
```

```
## Warning: package 'mvna' was built under R version 4.1.2
```

```
library(lattice) #Apoyo gráfico
```

```
## Warning: package 'lattice' was built under R version 4.1.2
```

```
library(cmprsk) #Paquetería enfocada a análisis de subdistribución de riesgos competitivos
```

```
## Warning: package 'cmprsk' was built under R version 4.1.2
```

```
## Loading required package: survival
```

```
library(etm) #Paquetería que nos permite estimar la matriz de transiciones de probabilidad
```

```
## Warning: package 'etm' was built under R version 4.1.2
```

```
#para modelos multiestado de tiempo no homogéneo de espacio finito, usando el  
#estimador Aalen-Johansen
```

Cargamos la base de datos

```
data(sir.adm)
```

Veamos las primeras 5 observaciones

```
head(sir.adm)
```

```
##      id pneu status time      age sex  
## 1    41     0      1    4 75.34153  F  
## 2   395     0      1   24 19.17380  M  
## 3   710     1      1   37 61.56568  M  
## 4  3138     0      1    8 57.88038  F  
## 5  3154     0      1    3 39.00639  M  
## 6  3178     0      1   24 70.27762  M
```

El dataset contiene una submuestra aleatoria de 747 pacientes y 6 variables:

id: Es un id generado aleatoriamente para cada paciente

pneu: Es un indicador de si el paciente presentaba neumonía (1) o no la presentaba (0) al momento de ser admitido en el estudio

status: Un indicador del estatus de la observación: 0 es una observación censurada, 1 es que el paciente fue dado de alta, 2 es que el paciente murió

time: Es el tiempo en días

age: La edad del paciente cuando entró al estudio

sex: F para mujer, M para hombre.

```
sum(is.na.data.frame(sir.adm))
```

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

Vemos que no hay ningún dato faltante (NA) en el dataset.

Estas 747 pacientes son de SIR 3 (Spread of nosocomial Infections and Resistant pathogens, es decir, Esparcimiento de infecciones nosocomiales y patógenos resistentes), un estudio conjunto en el hospital Charité de la Universidad de Berlin, Alemania, con una valoración prospectiva de datos para examinar el efecto de infecciones adquiridas en el hospital, en terapia-intensiva.

Notemos que se presenta una censura por la derecha.

El dataset contiene información en el estado de admisión de neumonía, tiempo de estadía en la unidad de terapia intensiva y “desenlace o resultado del tratamiento en unidad intensiva”, es decir, si se le dio de alta al paciente o este murió.

La neumonía es una infección severa, de la cual se tiene la sospecha, causa el requerimiento de cuidados adicionales (Es decir, una prolongada estadía en unidades de cuidado intensivo) y de incrementar la mortalidad.

```
sum(sir.adm$status==0)
```

```
## [1] 14
```

14 Observaciones censuradas, es decir, que seguían en unidad intensiva al final del estudio.

```
sum(sir.adm$status==1)
```

```
## [1] 657
```

657 Pacientes que fueron dados de alta

```
sum(sir.adm$status==2)
```

```
## [1] 76
```

76 pacientes fallecieron

```
sum(sir.adm$status==2 & sir.adm$pneu==1)
```

```
## [1] 21
```

21 de los pacientes que murieron tenían neumonía al momento de ser admitidos.

```
sum(sir.adm$status==2 & sir.adm$pneu==0)
```

```
## [1] 55
```

55 de los pacientes que murieron, no tenían neumonía al momento de ser admitidos

Esta base la elegimos porque nos parece un buen ejemplo de riesgos competitivos por lo siguiente:

Se investiga el tiempo hasta el final de la estadía y el estatus de el estado final: Si se dio de alta o murió en el hospital. Un desafío en el análisis de este dataset es que se ha encontrado que la neumonía incrementa la probabilidad de morir en el hospital, pero parece no tener efecto en el *riesgo* de muerte, es decir, no tiene

efecto en la probabilidad diaria de morir en el hospital, dado que uno sigue vivo y en la unidad de cuidados intensivos al inicio del día.

Los puntos finales competitivos son el ser dado de alta y morir en la unidad de terapia intensiva

## Análisis descriptivo y enfoque no paramétrico

### Función de riesgos acumulados

Necesitamos modificar primero el dataset en un dataset de tipo multiestado. Además, renombramos los datos para que, el evento de interés que es la muerte, corresponda al estado 1, y el estado 2 el evento competitivo:

```
#Transformamos sir.adm a un dataset de tipo multiestado
to <- ifelse(sir.adm$status == 0, "cens", ifelse(sir.adm$status == 1, 2, 1))
my.sir.data <- data.frame(id = sir.adm$id, from = 0, to, time = sir.adm$time, pneu = sir.adm$pneu)
```

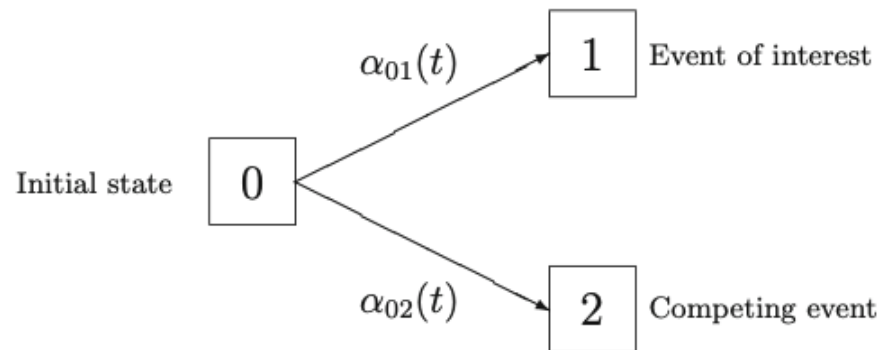
Notemos que my.sir.data tiene un componente pneu con el estatus de neumonía al momento de admisión. Revisamos que hicimos bien el renombramiento:

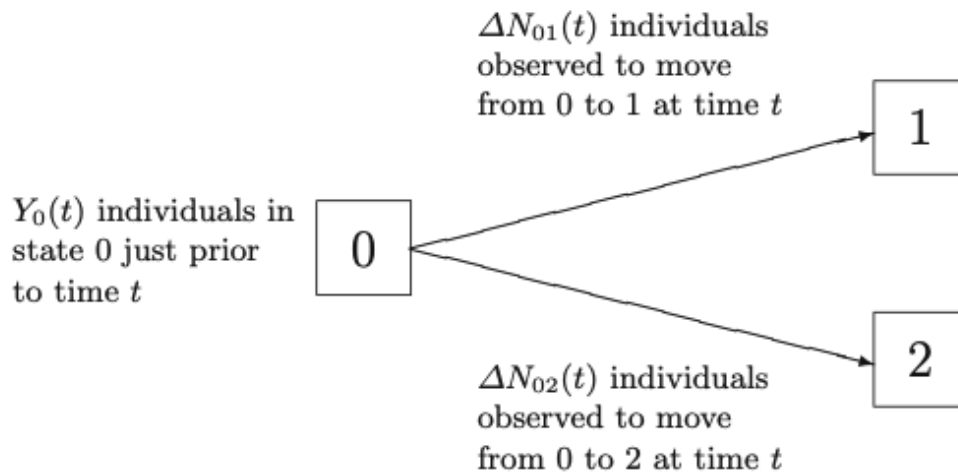
```
table(my.sir.data$to)
```

```
##
##      1      2 cens
##    76   657   14
```

Lo hicimos de manera correcta.

Antes de continuar, necesitamos describir el modelo de riesgos competitivos muliestado siguiente:





Lo hacemos definiendo una matriz de valores lógicos indicando los posibles tipos de transición en nuestro modelo multiestado:

```
tra <- matrix(FALSE, ncol = 3, nrow = 3)
dimnames(tra) <- list(c("0", "1", "2"), c("0", "1", "2"))
tra[1, 2:3] <- TRUE
tra
```

```
##      0      1      2
## 0 FALSE TRUE  TRUE
## 1 FALSE FALSE FALSE
## 2 FALSE FALSE FALSE
```

Esta matriz nos dice que un individuo se puede mover del estado 0 al estado 1, y del estado 0 al estado 2, pero las transiciones en el sentido contrario no son posibles; además los valores en diagonal están como falsos: Las transiciones de un estado a sí mismo no están modeladas. No es necesario un modelo para dicha “transición”. Los individuos que no hacen una transición a uno de los dos pares competitivos al tiempo  $t$ , permanece en el estado inicial 0 a tiempo  $t$ .

Ahora, calculamos la función de acumulación de riesgos específicos para muerte y dados de alta, respectivamente, y estratificados por su estatus de neumonía al momento de admisión:

```
## sin neumonía
my.nelaal.nop <- mvna(my.sir.data[my.sir.data$pneu == 0, ], c("0", "1", "2"), tra, "cens")
## con neumonía
my.nelaal.p <- mvna(my.sir.data[my.sir.data$pneu == 1, ], c("0", "1", "2"), tra, "cens")
```

Graficamos:

```
# Estilo de la gráfica
ltheme <- canonical.theme(color = FALSE)
ltheme$strip.background$col <- "white"
lattice.options(default.theme = ltheme)

#Plot para los que no tenían neumonía
dessin.nop <- xyplot(my.nelaal.nop, tr.choice = c("0 2", "0 1"), lwd = 2, layout = c(1, 2),
  strip = strip.custom( factor.levels = c("Sin neumonía:recuperación",
```

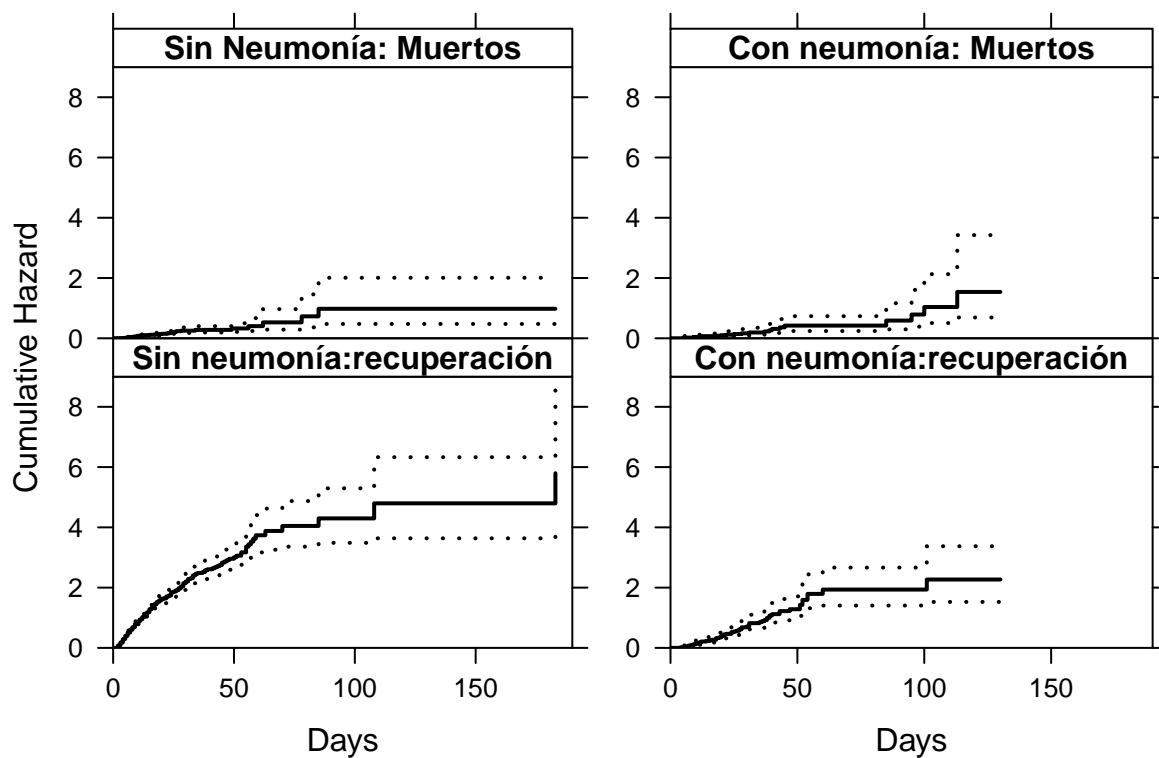
```

                                "Sin Neumonía: Muertos"),
                                par.strip.text = list(font = 2)),
ylim = c(0, 9), xlim = c(0, 190), xlab = "Days",
scales = list(alternating = 1, x = list(at = seq(0, 150, 50)),
              y = list(at = seq(0, 8, 2)))

#Plot para los que tenían neumonía
dessin.p <- xyplot(my.nelaal.p, tr.choice = c("0 2", "0 1"), lwd = 2, layout = c(1, 2),
                  strip = strip.custom( factor.levels = c("Con neumonía:recuperación",
                                                            "Con neumonía: Muertos"),
                                      par.strip.text = list(font = 2)), ylab = "",
                  ylim = c(0, 9), xlim = c(0, 190), xlab = "Days",
                  scales = list(alternating = 1, x = list(at = seq(0, 150, 50)),
                                y = list(at = seq(0, 8, 2))))

#Ploteamos ambas
print(dessin.nop, split = c(1, 1, 2, 1), more = TRUE, position = c(0, 0, 1.07, 1))
print(dessin.p, split = c(2, 1, 2, 1), position = c(-0.07, 0, 1, 1))

```



Esto no implica que la neumonía no tenga ningún efecto sobre la mortalidad. La razón es que la neumonía parece reducir el riesgo de ser dados de alta. Esto implica:

1. La neumonía parece reducir el riesgo por todas las causas al final de la de la unidad de cuidados intensivos.
2. Los pacientes con neumonía al ingreso permanecen más tiempo en la unidad. Durante esta estancia prolongada, están expuestos a un a un riesgo de muerte esencialmente

3. En consecuencia, mueren más pacientes con neumonía que sin neumonía.

Se trata de un fenómeno típico de riesgos competitivos. Como hay más de un riesgo que actúan sobre un individuo, no podemos saber, a partir de un solo riesgo, cuál será el curso futuro de un individuo. Esta situación se muestra de forma esquemática en la siguiente figura:

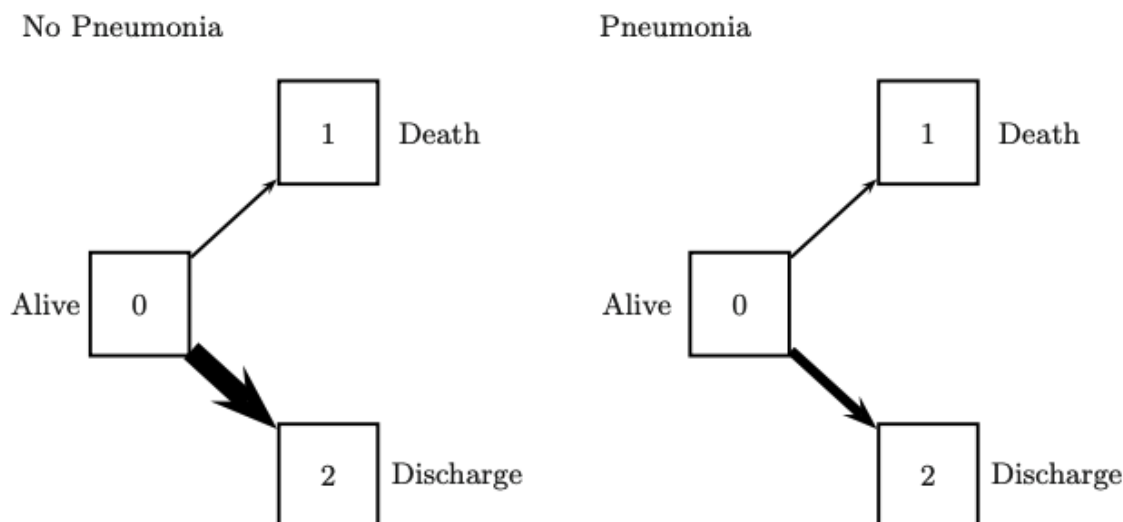


Fig. 1: Datos del hospital. Representación esquemática del efecto de la neumonía con causa específica. El estado de la neumonía no tiene ningún efecto sobre la causa específica de muerte, que es también un peligro menor. Un gráfico como el presente podían producirse con el paquete R `compeir`; hasta antes de ser deshabilitada del CRAN

Recordemos que una forma de pensar en los peligros de causas específicas es en términos de fuerzas momentáneas de transición que se mueven a lo largo de las flechas de los cuadros multiestado. La magnitud de estas fuerzas se muestra de forma esquemática en la figura 1. La “fuerza de muerte” no está influenciada por el estado de la neumonía, pero la “fuerza del alta” se reduce sustancialmente de la neumonía en el momento del ingreso. La figura 1 ilustra que la “fuerza global”, es decir, el riesgo por todas las causas que arrastra un individuo se ve reducido, conduciendo a una mayor permanencia en la unidad, y que la fuerza relativa entre las fuerzas específicas de la causa de muerte y de alta, se ve modificada por el estado de la neumonía.

Notemos que la representación esquemática de la figura 1 tiene limitaciones. La magnitud de las fuerzas de transición momentánea no suele ser constante sobre el tiempo, de modo que necesitaríamos toda una serie de gráficos como los de la figura 1. De hecho, esto se consigue en la gráfica que ploteamos anteriormente de los estimadores de Nelson-Aalen; la forma de los estimadores de dichos estimadores, que estiman los peligros acumulativos, está determinada por los peligros específicos de la causa. También podemos pensar en la figura 1 de una manera que no necesariamente ilustre la magnitud de los peligros, pudiendo variar con el tiempo, sino únicamente los cocientes de los peligros de muerte y los cocientes de los peligros de descarga, respectivamente, suponiéndose constantes. Este es el enfoque adoptado por la modelización de los riesgos proporcionales a la causa.

## Función de incidencia acumulada

Por último, comprobamos si nuestra interpretación del análisis de riesgos acumulativos ha sido correcta observando los estimadores de Aalen-Johansen de las funciones de incidencia acumulada, nuevamente estratificadas por el estado de la neumonía. Recordemos que la función de incidencia acumulada para la muerte, por ejemplo, muestra la proporción esperada de individuos que mueren en la unidad a lo largo del

tiempo. Si nuestra interpretación del análisis de riesgos acumulativos ha sido correcta, la función de incidencia acumulativa estimada para la muerte..... dentro de los pacientes con neumonía, debería estar por encima de los pacientes sin neumonía.

Utilizando la función `cuminc` del paquete `cmprsk`, calculamos las estimaciones  $\mathbb{P}(T \leq t, X_T = j), j = 1, 2$  (Es decir, de la función acumulativa de incidencia) dentro de los grupos definidos

```
my.sir.cif <- cuminc(my.sir.data$time, my.sir.data$to, group=my.sir.data$pneu, cencode="cens")
my.sir.cif
```

```
## Tests:
##      stat      pv df
## 1 16.16525 5.804939e-05 1
## 2 42.07480 8.784784e-11 1
## Estimates and Variances:
## $est
##      50      100      150
## 0 1 0.07930062 0.08576614 0.08576614
## 1 1 0.18393349 0.22550113      NA
## 0 2 0.88998816 0.91100110 0.91261748
## 1 2 0.64633198 0.73293123      NA
##
## $var
##      50      100      150
## 0 1 0.0001125301 0.0001222882 0.0001222882
## 1 1 0.0016495598 0.0020723886      NA
## 0 2 0.0001517978 0.0001266104 0.0001250158
## 1 2 0.0025299050 0.0022581560      NA
```

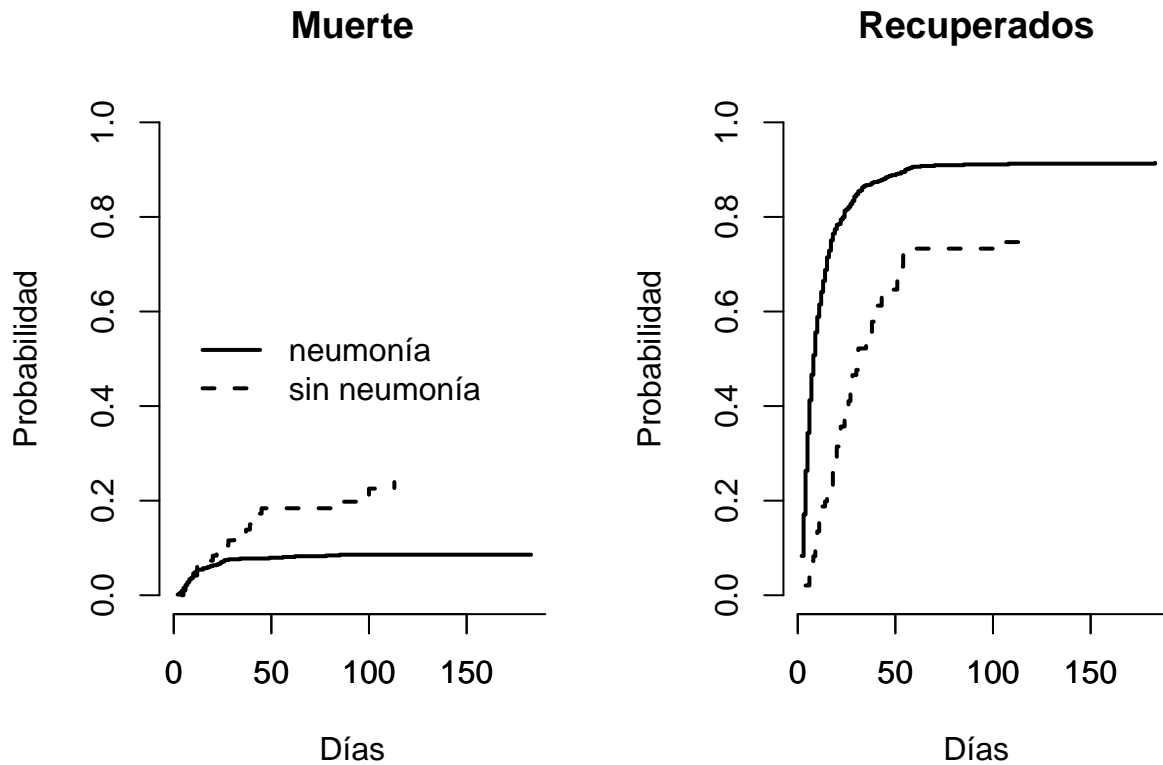
El valor regresado por `cuminc` es una lista con los componentes “0 1”, “1 1”, “0 2” y “1 2”. Los componentes “0 1” y “1 1” contienen resultados para el tipo de fallo 1 (Muerte); los componentes “0 2” y “1 2” contienen resultados para el tipo de fallo 2 (Recuperación). Los componentes “0 1” y “0 2” son para pacientes con estado de neumonía 0 al ingreso, es decir, sin neumonía, y los componentes “1 1” y “1 2” son para pacientes con estado de neumonía 1.

Esto también lo podemos hacer con la paquetería `etm`, para matrices de transiciones; al igual que con `mvna`. Ejecutamos `etm` con cada estrato:

```
my.sir.etm.nop <- etm(my.sir.data[my.sir.data$pneu == 0, ], c("0", "1", "2"), tra, "cens", s = 0)
my.sir.etm.p <- etm(my.sir.data[my.sir.data$pneu == 1, ], c("0", "1", "2"), tra, "cens", s = 0)
```

Graficando:

```
op <- par(mfrow = c(1, 2))
#Muerte
plot(my.sir.etm.nop, tr.choice = "0 1", conf.int = FALSE, lwd = 2, lty = 1, xlab = "Días",
     ylab = "Probabilidad", bty = "n", legend = FALSE)
lines(my.sir.etm.p, tr.choice = "0 1", conf.int = FALSE, lwd = 2, lty = 2)
legend(0, 0.6, c("neumonía", "sin neumonía"), col = 1, lty = c(1, 2), bty = "n", lwd = 2)
title("Muerte")
axis(1, at = seq(0, 200, 50))
##Datos de alta
plot(my.sir.etm.nop, tr.choice = "0 2", conf.int = FALSE, lwd = 2, lty = 1, xlab = "Días",
     ylab = "Probabilidad", bty = "n", legend = FALSE)
lines(my.sir.etm.p, tr.choice = "0 2", conf.int = FALSE, lwd = 2, lty = 2)
axis(1, at = seq(0, 200, 50))
title("Recuperados")
```



```
par(op)
```

Esta gráfica presenta las estimaciones de Aalen-Johansen  $\mathbb{P}(bT \leq t, X_T = j)$  de las funciones de incidencia acumulativa para la muerte (izquierda,  $j = 1$ ) y para el alta (derecha,  $j = 2$ ), estratificadas por el estado de la neumonía al ingreso. Las líneas continuas corresponden a pacientes sin neumonía.

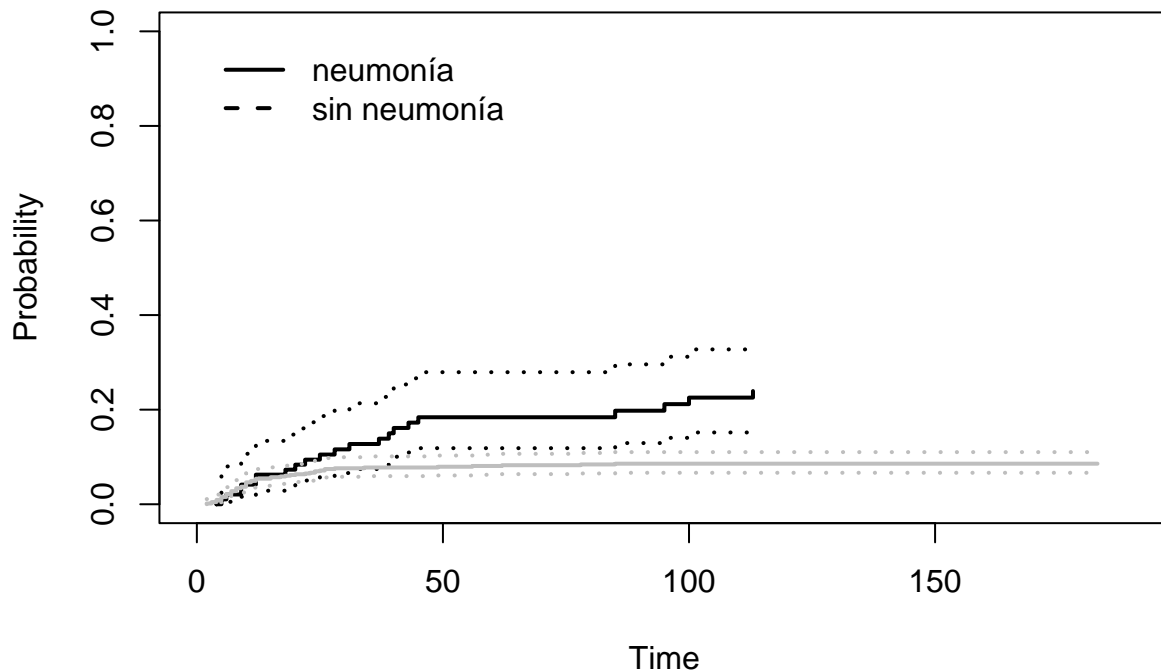
Como era de esperarse, encontramos que mueren más pacientes entre los que tienen neumonía.

Las estimaciones de Aalen-Johansen  $\mathbb{P}(T \leq t, X_T = 1)$  se muestran en la siguiente gráfica, junto con intervalos de confianza del 95% puntuales, generadas por:

```
plot(my.sir.etm.p, tr.choice = '0 1', col = 1, lwd = 2, conf.int = TRUE, ci.fun = "cloglog",
     legend = FALSE, ylab="Probability", xlim=c(0,190))
lines(my.sir.etm.nop, tr.choice = '0 1', col = "gray", lwd = 2, conf.int = TRUE,
     ci.fun = "cloglog")
legend(0, 1, c("neumonía", "sin neumonía"), col = 1, lty = c(1, 2), bty = "n", lwd = 2)
title("Estimador Aalen-Johansen con intervalos de confianza")
```



## Estimador Aalen–Johansen con intervalos de confianza



Los intervalos de confianza apoyan nuestra conclusión anterior de que finalmente vemos más casos de muerte en el grupo de pacientes con neumonía. Los gráficos de riesgos acumulados, como en la fgráfica de riesgos acumulados (La primera gráfica presentada), y los gráficos de funciones de incidencia acumulada, como en las Figuras 2 y 3, ambos tienen sus méritos relativos: Obviamente, en la figura 3 es más fácil saber si la neumonía aumenta la mortalidad unitaria. Sin embargo, tenemos que examinar los peligros acumulados por causas específicas para ver si el aumento de la mortalidad se debe a un aumento del peligro de muerte, o, como en el presente ejemplo, a una disminución del riesgo de muerte.

## Modelo de riesgos proporcionales de causa específica

En el apartado anterior dimos un análisis no paramétrico.

Recordemos, en particular, que a partir de las estimaciones de Nelson-Aalen de los riesgos acumulativos por causas específicas en la primera gráfica que hicimos, que la neumonía de ingreso aumenta la mortalidad hospitalaria a través de un efecto decreciente sobre el riesgo de alta en vida, mientras que el riesgo de muerte en el hospital se mantiene esencialmente sin cambios. El objetivo del presente análisis es volver a investigar este hallazgo mediante modelos de riesgos proporcionales a la causa. Utilizamos el data frame `my.sir.data` generado al inicio. La primer gráfica que hicimos sugiere que la neumonía tiene diferentes efectos en los riesgos específicos de la causa. Por lo tanto, simplemente ajustamos dos modelos de Cox diferente. Este es el resultado del análisis de la causa específica de interés para la muerte en el hospital,

```
fit.pneu.01 <- coxph(Surv(time, to == 1) ~ pneu, my.sir.data)
fit.pneu.02 <- coxph(Surv(time, to == 2) ~ pneu, my.sir.data)
summary(fit.pneu.02)
```

```
## Call:
```

```
## coxph(formula = Surv(time, to == 2) ~ pneu, data = my.sir.data)
```

```
##
##   n= 747, number of events= 657
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## pneu -1.0901    0.3362   0.1299 -8.391   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## pneu    0.3362      2.974    0.2606    0.4337
##
## Concordance= 0.577 (se = 0.008 )
## Likelihood ratio test= 91.7  on 1 df,   p=<2e-16
## Wald test               = 70.4  on 1 df,   p=<2e-16
## Score (logrank) test = 77.09  on 1 df,   p=<2e-16
summary(fit.pneu.02)
```

```
## Call:
## coxph(formula = Surv(time, to == 2) ~ pneu, data = my.sir.data)
##
##   n= 747, number of events= 657
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## pneu -1.0901    0.3362   0.1299 -8.391   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## pneu    0.3362      2.974    0.2606    0.4337
##
## Concordance= 0.577 (se = 0.008 )
## Likelihood ratio test= 91.7  on 1 df,   p=<2e-16
## Wald test               = 70.4  on 1 df,   p=<2e-16
## Score (logrank) test = 77.09  on 1 df,   p=<2e-16
```

Los análisis de riesgos proporcionales específicos a la causa concuerdan con nuestros resultados anteriores. El resultado de R anterior también vuelve a poner de relieve dos aspectos importantes en el análisis de los datos de riesgos concurrentes: en primer lugar, todos los riesgos específicos de las causas deben analizarse. No debemos concluir de ninguna manera a partir de una razón de riesgo de muerte específica por causa de 0,85 con un intervalo de confianza del 95% [0,503, 1,437] que la neumonía parece no tener impacto en la muerte hospitalaria. En segundo lugar, los cocientes de riesgo específicos de la causa que se muestran arriba, no hacen ninguna declaración sobre la magnitud de los riesgos base específicos de la causa. Esto es muy diferente a nuestro análisis inicial basado en los estimadores de Nelson-Aalen en la gráfica inicial.

```
a01.0 <- basehaz(fit.pneu.01, centered=FALSE)
a02.0 <- basehaz(fit.pneu.02, centered=FALSE)

split.screen(figs=c(1,2))
```

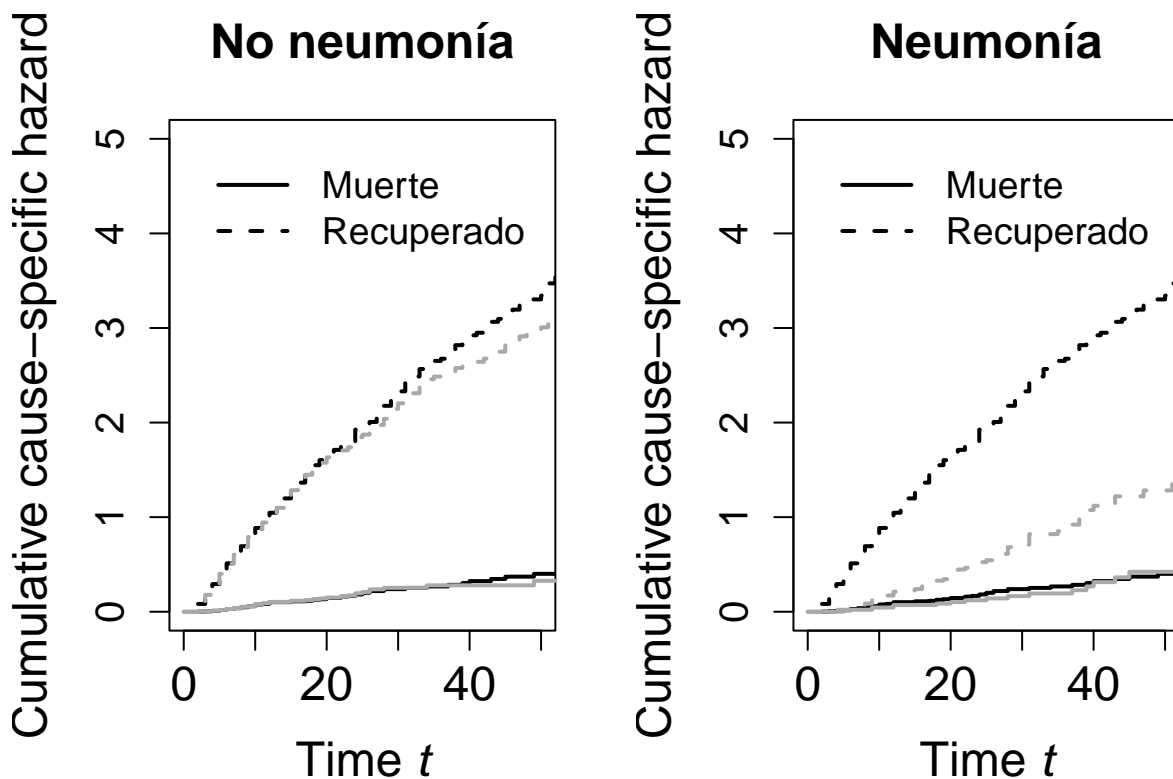
```
## [1] 1 2

screen(1)
plot(c(0, 50), c(0, 5), xlab = expression(paste(Time, " ", italic(t))),
     ylab = "Cumulative cause-specific hazard", type = "n", axes = FALSE,
     main = "No neumonía", cex.main = 1.5, cex.lab = 1.5)
```

```

axis(1, at=seq(0, 50, 10), cex.axis=1.5)
axis(2, at=seq(0, 5, 1), cex.axis=1.25)
box()
lines(a02.0$time, a02.0$hazard, type="s", lwd=2, lty=2)
lines(a01.0$time, a01.0$hazard, type="s", lwd=2)
lines(my.nelaal.nop, conf.int = FALSE, col = rep("darkgray", 2), lty = c(1, 2), lwd = 2)
legend(0,5,c("Muerte", "Recuperado"), lty=1:2,bty="n", cex=1.2, lwd=2)
screen(2)
plot(x=c(0, 50), y=c(0, 5), xlab=expression(paste(Time, " ", italic(t))),
     ylab="Cumulative cause-specific hazard", type="n", axes=F,
     main="Neumonía", cex.main=1.5, cex.lab=1.5)
axis(1, at=seq(0, 50, 10), cex.axis=1.5)
axis(2, at=seq(0, 5, 1), cex.axis=1.25)
box()
lines(a02.0$time, a02.0$hazard, type="s", lwd=2, lty=2)
lines(a01.0$time, a01.0$hazard, type="s", lwd=2)
lines(my.nelaal.p, conf.int = FALSE, col = rep("darkgray", 2), lty = c(1, 2), lwd = 2)
legend(0,5,c("Muerte", "Recuperado"), lty=1:2,bty="n", cex=1.2, lwd=2)

```



```
close.screen(all.screens=TRUE)
```

La gráfica anterior es de los estimadores de Nelson-Aalen junto con los estimadores de Breslow (para la ausencia de neumonía) y los estimadores de riesgo acumulado (para la neumonía al ingreso). Encontramos que todas las curvas específicas de la muerte en el hospital están en buena concordancia, al igual que los estimadores de referencia del riesgo acumulado en el alta. Sin embargo, los estimadores respectivos del riesgo acumulado en el alta para los pacientes con neumonía no coinciden tanto, indicando que el efecto de la

neumonía en el riesgo en el alta puede no seguir un modelo de riesgos proporcionales a la causa.

La gráfica sugiere que la razón de riesgo de alta estimada por causa específica es de 0,336 con un intervalo de confianza del 95% [0,261, 0,434] (Es decir, los datos del summary) informa de un efecto promediado al tiempo de la neumonía sobre el riesgo de alta.

## Modelo de subdistribución de riesgos proporcionales

Consideramos el modelo de subdistribución

```
fit.sir<-crr(ftime = my.sir.data$time, fstatus = my.sir.data$to, cov1 = my.sir.data$pneu,
            failcode = "1", cencode = "cens")
fit.sir
```

```
## convergence: TRUE
## coefficients:
## my.sir.data$pneu1
##           0.9749
## standard errors:
## [1] 0.2496
## two-sided p-values:
## my.sir.data$pneu1
##           9.4e-05
```

El valor reportado, es el p-value correspondiente a un test del tipo log-rank para la subdistribución de riesgos (Abarcaremos esto más a detalle al final, en la conclusión)

Veamos un summary:

```
summary(fit.sir)

## Competing Risks Regression
##
## Call:
## crr(ftime = my.sir.data$time, fstatus = my.sir.data$to, cov1 = my.sir.data$pneu,
##      failcode = "1", cencode = "cens")
##
##              coef exp(coef) se(coef)      z p-value
## my.sir.data$pneu1 0.975      2.65      0.25 3.91 9.4e-05
##
##              exp(coef) exp(-coef) 2.5% 97.5%
## my.sir.data$pneu1      2.65      0.377 1.63  4.32
##
## Num. cases = 747
## Pseudo Log-likelihood = -491
## Pseudo likelihood ratio test = 12.2 on 1 df,
```

Vemos que el coeficiente es significativo: Encontramos un efecto significativo de la neumonía en la función de incidencia acumulada por muerte hospitalaria 2.65 veces mayor (usando el estimador puntual), puesto que nos da un intervalo de confianza para el coeficiente de [1.63, 4.32]; como el coeficiente es positivo, en el contexto de riesgos competitivos, nos indica que la neumonía incrementa la mortalidad en los hospitales (Es decir, de las personas que se encuentran ya hospitalizadas), sin embargo, no muestra algún efecto en la causa específica para muerte de pacientes hospitalizados. El incremento en la mortalidad fue debido a un considerable decrecimiento del riesgo de causa específica para vivos dados de alta. (Todo esto porque la base es que no ingresó con neumonía)

Ahora, usemos dicho modelo para modelar la CIF:

```

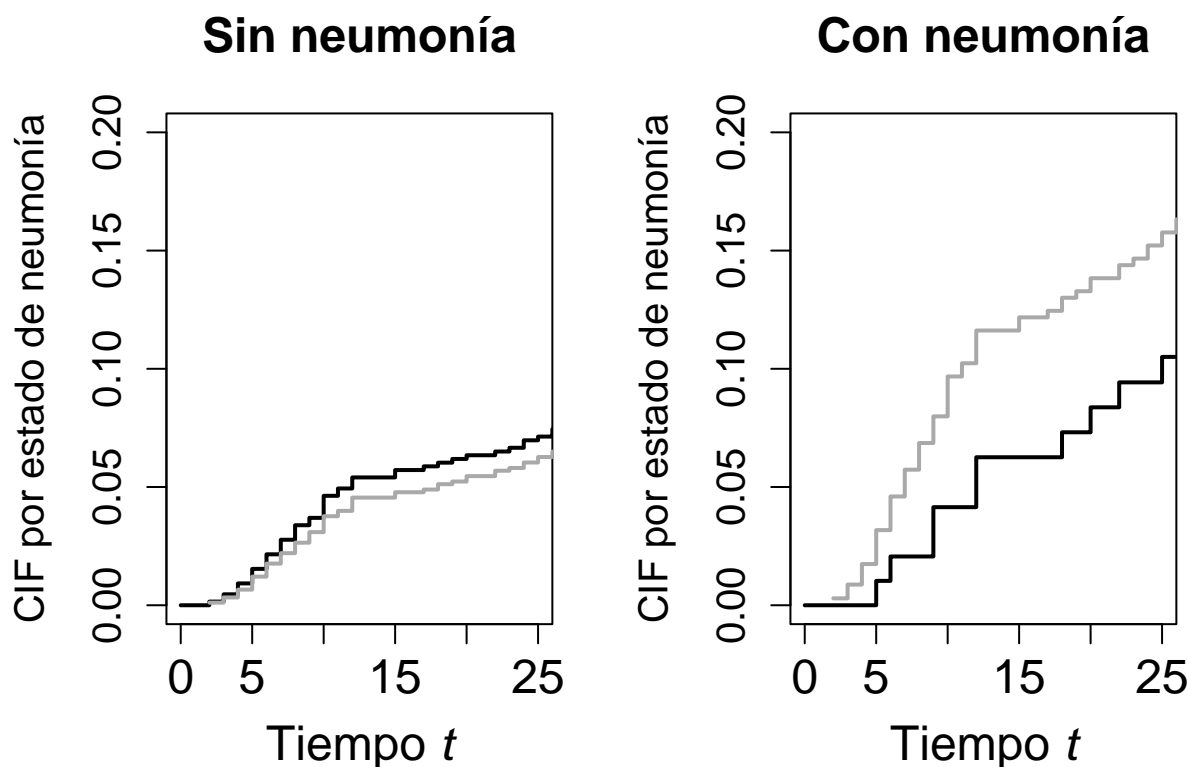
daddeln <- predict.crr(fit.sir, cov1 = matrix(c(0, 1), nrow = 2))

## predicción sin neumonía
split.screen(figs=c(1,2))

## [1] 1 2

screen(1)
plot(c(0, 25), c(0, 0.2), xlab = expression(paste(Tiempo, " ", italic(t))),
     ylab = "", type = "n", axes=FALSE, main = "Sin neumonía", cex.main = 1.5, cex.lab = 1.5)
axis(1, at = seq(0, 25, 5), cex.axis = 1.5)
axis(2, at = seq(0, 0.2, 0.05), cex.axis = 1.25)
box()
mtext(text="CIF por estado de neumonía", side = 2, line = 3, cex = 1.25)
lines(my.sir.cif$`0 1`$time, my.sir.cif$`0 1`$est, type = "s", lwd = 2, lty = 1)
## predicción con neumonía
lines(daddeln[,1], daddeln[,2], type = "s", lwd = 2, lty = 1, col = "darkgrey")
screen(2)
plot(c(0, 25), c(0, 0.2), xlab = expression(paste(Tiempo, " ", italic(t))),
     ylab = "", type = "n", axes = FALSE, main = "Con neumonía", cex.main = 1.5, cex.lab = 1.5)
axis(1, at = seq(0, 25, 5), cex.axis = 1.5)
axis(2, at = seq(0, 0.2, 0.05), cex.axis = 1.25)
box()
mtext(text="CIF por estado de neumonía", side = 2, line = 3, cex = 1.25)
lines(my.sir.cif$`1 1`$time, my.sir.cif$`1 1`$est, type = "s", lwd = 2, lty = 1)
## predicted curves
lines(daddeln[,1], daddeln[,3], type = "s", lwd = 2, col = "darkgrey")

```



```
close.screen(all.screens=TRUE)
```

Las líneas negras están dadas por el estimador de Aalen-Johansen, y las grises por el modelo bajo el supuesto de modelo de subdistribución de riesgos proporcionales; vemos para las personas sin neumonía es prácticamente la misma, pero no es el caso para las personas con neumonía inicialmente.

## Entonces, ¿Sí influye el que ingrese con neumonía o sin neumonía?

Revisamos brevemente el ejemplo de datos hospitalarios, donde el objetivo era investigar el impacto de neumonía diagnosticada al ingreso. El resultado no fue una diferencia significativa entre los peligros específicos de la causa de muerte en el hospital, pero los peligros específicos de la causa significativamente diferentes para el alta con vida. Este ejemplo, se ilustran las limitaciones de probar la igualdad de riesgos de causa específica. Nuestro análisis anterior mostró que la neumonía aumentó el número de pacientes que mueren en el hospital, aunque el riesgo para muerte hospitalaria se encontrado que es similar con o sin neumonía.

El hallazgo encontrado en nuestro análisis ha sido que una interpretación adecuada de un análisis de riesgos competitivos requiere una consideración cuidadosa de todos los riesgos específicos de la causa, incluidos los signos y las magnitudes de los efectos de, por ejemplo, el estado de neumonía así como la magnitud relativa de los riesgos específicos de causa única dentro de un grupo.

El modelo anterior se usa a menudo para estudiar directamente el impacto de covariables como la neumonía, Este modelo es un modelo de tipo Cox para el riesgo de subdistribución, y que la subdistribución el peligro restablece una correspondencia biyectiva con la función de incidencia acumulada de interés. Por lo tanto, el p-value obtenido al ajustar tal modelo debe ser adecuado para comparar funciones de incidencia acumulada directamente. Ilustramos esto brevemente con los datos del hospital.

```
fit.sir
```

```
## convergence: TRUE
## coefficients:
## my.sir.data$pneu1
##           0.9749
## standard errors:
## [1] 0.2496
## two-sided p-values:
## my.sir.data$pneu1
##           9.4e-05
```

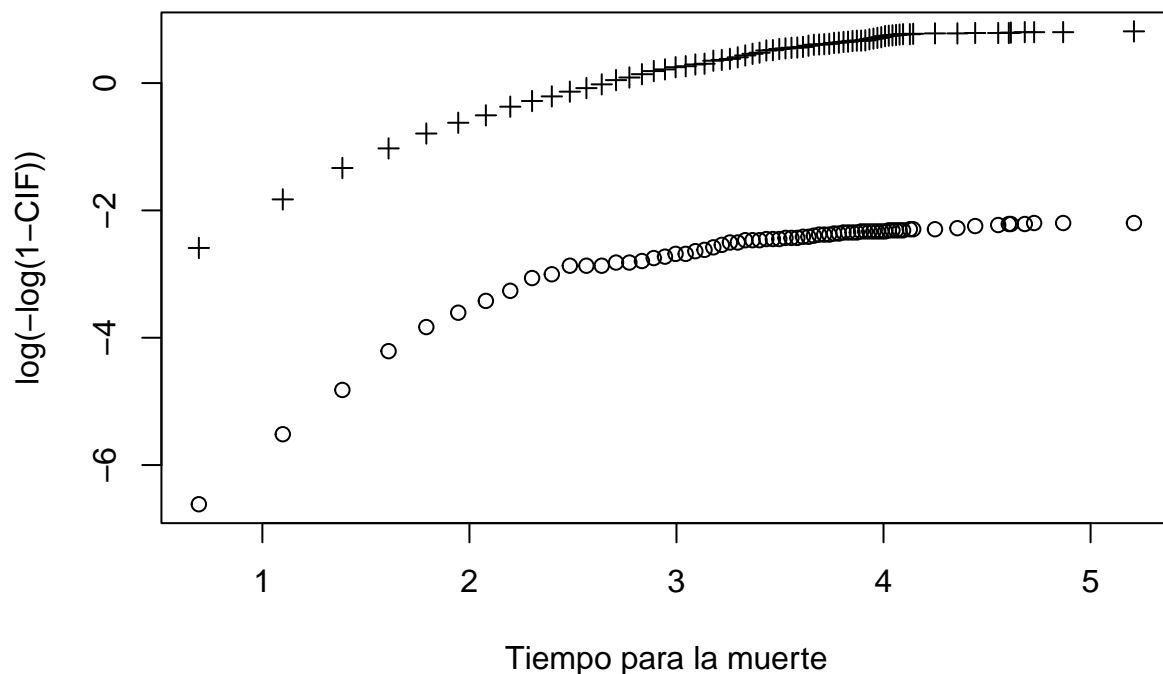
Encontramos un efecto significativo de la neumonía en la función de incidencia acumulada por muerte hospitalaria. Esta prueba se debe a Gray y, por lo tanto, a menudo se llama Prueba de Gray, y consiste en un test no paramétrico en el que se comparan dos o más CIFs. El test es análogo al log-rank para comparar curvas derivadas de los estimadores de Kaplan-Meier, usando una estadística de prueba modificada de la ji-cuadrada. La prueba puede derivarse en analogía con nuestra derivación de la prueba log-rank ya vista en clase, si usamos los incrementos del estimador de la función de riesgo de subdistribución acumulativa.

## Comprobación de supuestos

Se necesita, en riesgos proporcionales *en la CIF*.

Para comprobar la hipótesis de proporcionalidad de la regresión de riesgos competitivos, podemos plotear  $\ln(-\ln(1 - F))$  contra  $\ln(\text{tiempo})$ , donde F es la CIF del evento de interés.

```
fit=cuminc(my.sir.data$time,my.sir.data$to,cencode = "cens",strata=my.sir.data$pneu)
a=timepoints(fit,times=my.sir.data$time)
cif=t(a$est[1:2,])
llcif=log(-log(1-cif))
matplot(log(unique(sort(my.sir.data$time))),llcif,pch=c(1,3),
col=1,
xlab='Tiempo para la muerte', ylab='log(-log(1-CIF))')
```



Notamos que las curvas difieren a una distancia casi similar durante todo el tiempo, lo que prueba que dichos riesgos son proporcionales

En el modelado de riesgos de causa específica, se asume que el logaritmo del riesgo cambia linealmente con la covariable. Puede ser revisada categorizando la covariable y examinando los efectos para cada categoría; sin embargo nuestras variables de interés ya son categóricas, y el test de Gray nos garantiza que hay diferencia entre estas; por lo que se cumple la linealidad.

A manera de resumen, recordemos que, a mediados del siglo XVII John Graunt, un comerciante de mercería de oficio, empezó a recopilar y tabular la información incluida en las Cartas de Mortalidad publicadas semanalmente en ese tiempo en Londres. Este trabajo, probablemente realizado en colaboración con Sir William Petty, apareció en 1662 y contenía un resumen demográfico de las causas de muerte en Inglaterra y Gales. Como resultado, John Graunt fue elegido miembro de la Royal Society con recomendación del rey Carlos II, hazaña nada despreciable teniendo en cuenta la estructura jerárquica de la sociedad existente en ese momento.

A partir del siglo XVIII, se desarrollaron las matemáticas, se cambiaron las notaciones y se ramificó la estadística como una ciencia por derecho propio. Los grandes descubrimientos del siglo XIX en la física y la biología condujo a una vista determinista del universo. Sin embargo, a medida que crecieron las herramientas de medición cada vez más se hizo evidente que algunos inexplicables, factores aleatorios estaban en el trabajo.

Veamos que, para entender que es un riesgo competitivo, vemos el análisis de datos de “tiempo para un evento”, los cuales consisten de pares de observaciones para cada individuo (el primero es la longitud de tiempo durante el cual ningún evento fue observado; el segundo es un indicador de si el final de dicho tiempo corresponde a un evento o solo al final de la observación), se suele referir como análisis de supervivencia, porque la teoría se desarrolló utilizando la muerte como evento de interés. También se le llama análisis de datos incompletos, donde este término de incompletos, se refiere a que un evento no ha sido observado, pero está obligado a tener lugar si se sigue por el tiempo suficiente. En presencia de riesgos competitivos, los métodos habituales de supervivencia deben aplicarse con precaución y hay que ser conscientes de la



consecuencias de su uso. El método Kaplan-Meier es el más común, así como la técnica más controvertida en el marco de riesgos competitivos. Es un método para estimar probabilidades de supervivencia (Kaplan y Meier, 1958) en diferentes momentos, como hemos visto en clase. Es relativamente fácil de aplicar e interpretar y se puede representar visualmente.

Viendo desde un enfoque más práctico, la situación en la que están presentes los riesgos competitivos ha sido expresada de diferentes maneras por diferentes autores. Dado que muchos de estas descripciones arrojan luz sobre diferentes aspectos de la situación de los riesgos competitivos, varios se presentan aquí para ilustrar mejor el punto. Kalbfleisch y Prentice (2002) describieron los riesgos competitivos como la situación en la que un individuo puede experimentar más de un tipo de evento. Por otro lado, las explicaciones más generales y ejemplos de otros campos de la investigación médica también son introducidos. En la investigación del cáncer, los pacientes se someten a uno o más de los tres principales tipos de tratamiento: cirugía, quimioterapia y radiación terapéutica. Supongamos que el tratamiento parece exitoso y todos se eliminó la evidencia de la enfermedad. Un punto final común de interés en estudios de cáncer es el tiempo hasta el regreso de la enfermedad (recurrencia) después del aparente éxito inicial del tratamiento.

Hay dos enfoques matemáticos para hacer frente a la competencia de riesgos, y ambos han recibido atención en la literatura. En esta sección las dos definiciones matemáticas de riesgos competitivos se introducen, así como los conceptos principales: subdistribución, subdensidad, subpeligro, peligro de causa específica y el peligro de la subdistribución. En ausencia de riesgos competitivos, los datos de supervivencia suelen ser presentados como una variable aleatoria bivariada o par  $(T, C)$ . La variable de censura,  $C$ , es 1 si se observó el evento de interés, y es 0 si la observación fue censurada. Cuando  $C = 1$  el primer miembro del par,  $T$ , es el momento en que ocurrió el evento y cuando  $C = 0$ ,  $T$  es el momento en que se censuró la observación.

La segunda definición matemática del resultado de riesgos competitivos se escribe en términos de tiempos de eventos latentes o no observados  $T_1, T_2, \dots, T_p$ , uno para cada uno de los tipos de eventos  $p$ , en la competencia situación de riesgos sólo se observa el primer evento, por lo que la variable de tiempo,  $T$ , se define como la primera de estos hipotéticos tiempos no observados. El CIF es la probabilidad de que un evento de tipo  $i$  ocurre en o antes del tiempo  $t$ . La función de distribución general es la probabilidad de que un evento de cualquier tipo ocurre en o antes del tiempo  $t$ .

Con las curvas de Kaplan-Meier, aproximamos la función de supervivencia, expresando la probabilidad de que el evento de interés (por ejemplo, el evento de muerte) no ocurra en un cierto tiempo  $t$ . Un enfoque alternativo para visualizar la información agregada de un conjunto de datos centrado en la supervivencia implica el uso de la función de riesgo, que se puede interpretar como la probabilidad de que el sujeto experimente el evento de interés dentro de un pequeño intervalo de tiempo, asumiendo que el sujeto ha sobrevivido hasta el comienzo de dicho intervalo.

En cuanto al modelo multiestado de riesgos competitivos vemos que, inicialmente cada individuo está en el estado 0 en el origen del tiempo. El individuo permanece en este estado hasta que ocurra algún primer evento. Por lo general, hay un evento de interés, modelado por transiciones al estado 1, y todos los demás primeros tipos de eventos son incluidos en el estado 2 de eventos competitivos.

En cuanto a los modelos proporcionales de causa específica, primero consideramos la estimación de los coeficientes de regresión. A continuación, consideramos la estimación basada en modelos de los riesgos de línea de base específicos de la causa acumulada y, finalmente, se introduce la predicción de las funciones de incidencia acumulada.

El estimador de tipo Nelson-Aalen, a menudo se denomina estimador de Breslow (Breslow, 1972). En análisis de riesgos proporcionales estándar (punto final único) datos de supervivencia, el riesgo de línea de base por todas las causas con frecuencia no se considera, ya que se cancela de la verosimilitud parcial. Esto es conveniente en que el analista de datos no necesita considerar el riesgo de línea de base para estudiar el impacto de las covariables en la supervivencia.

En cuanto al modelo de riesgos de subdistribución proporcional vemos que, el análisis de los datos de riesgos competitivos requiere analizar todos los factores específicos de la causa de riesgos. Como se explica en la introducción, el objetivo de un análisis de subdistribución de riesgos es proporcionar un análisis único de una cantidad, el riesgo de subdistribución, que permite la interpretación directa de los resultados en términos de

una función de incidencia acumulada, tal como podríamos interpretar el análisis único de un riesgo por todas las causas en la sección anterior en términos de la distribución del tiempo de espera en el estado inicial.

En cuanto al modelado en presencia de riesgos competitivos, se presenta el trabajo de Fine y Gray (1999) sobre modelado de la peligrosidad del CIF. La metodología tiene muchas puntos en común con la prueba de k-muestra introducida por Gray (1988) y el modelo de riesgos proporcionales de Cox. Aunque la teoría se desarrolla gradualmente (a partir del caso de no censura), se presenta el caso general de datos incompletos aquí. Este es el tipo de datos que solemos encontrar: hay personas físicas con un evento de interés registrado, otras con tipo de evento de riesgo competitivo y otros que están censurados.

El hecho de que el análisis de subdistribución permita una interpretación de probabilidad directa lo ha hecho popular en las aplicaciones. Esto es particularmente relevante, porque en los análisis de riesgos competitivos aplicados basados en los peligros habituales de causa específica, a menudo falta un análisis del peligro competitivo de causa específica. Una limitación del enfoque de riesgos de subdistribución en comparación con los riesgos de causa específica es que solo estos últimos determinan completamente el comportamiento estocástico de todo el proceso de riesgos competitivos. El riesgo de subdistribución solo especifica la función de incidencia acumulada de interés, pero no las funciones de incidencia acumulada competidoras. Conceptualmente, este enfoque presenta problemas si se supone que ambos riesgos de subdistribución siguen un modelo de riesgos de subdistribución proporcional.

De igual manera, se les invita a leer el archivo adjuntado que lleva por nombre [Proyecto Riesgos Competitivos TrabajoEscrito], ya que, ahí podrán encontrar de una manera más completa y general toda la información que se intento explicar a través de documento.