

# Fontecha - TP 8 Tiempo hasta un evento

María Eugenia Fontecha

09 noviembre, 2020

El archivo “Dieta de ratas.xls” contiene el tiempo libre de tumor (en días) de 90 ratas con 3 dietas diferentes. Se dividieron a 90 ratas de laboratorio en tres grupos que fueron alimentadas con dietas “low-fat”, “saturated” y “unsaturated” respectivamente. Las ratas tenían la misma edad y eran de la misma especie y eran similares en cuanto a su condición física. Se inyectó una idéntica cantidad de células tumorales en la pata de cada rata. Las ratas se observaron por 200 días. Muchas desarrollaron tumores reconocibles tempranamente en el estudio, algunas no desarrollaron tumor al término del estudio. La rata 16 del grupo 1 y la 24 del tercero se murieron accidentalmente luego de 140 y 170 días respectivamente, sin evidencia de tumor. La tabla indica para cada rata en cada dieta, el tiempo libre de tumor en días: el tiempo desde la inyección hasta el tiempo en el que el tumor se desarrolla o el fin del estudio. Todas las ratas del grupo “unsaturated” desarrollaron el tumor antes de los 200 días.

```
dieta <- read.xlsx('Dieta de ratas.xls', 'Hoja1', header=TRUE)
summary(dieta)
```

##	Rat	Group	time	Censored
##	Min. : 1.0	Length:90	Min. : 43.0	Min. :0.0000
##	1st Qu.: 8.0	Class :character	1st Qu.: 77.0	1st Qu.:1.0000
##	Median :15.5	Mode :character	Median :109.5	Median :1.0000
##	Mean :15.5		Mean :123.3	Mean :0.7556
##	3rd Qu.:23.0		3rd Qu.:177.8	3rd Qu.:1.0000
##	Max. :30.0		Max. :200.0	Max. :1.0000

```
dieta$Group <- as.factor(dieta$Group)
str(dieta$Group)
```

```
## Factor w/ 3 levels "Lowfat","Saturated",...: 1 1 1 1 1 1 1 1 1 1 ...
```

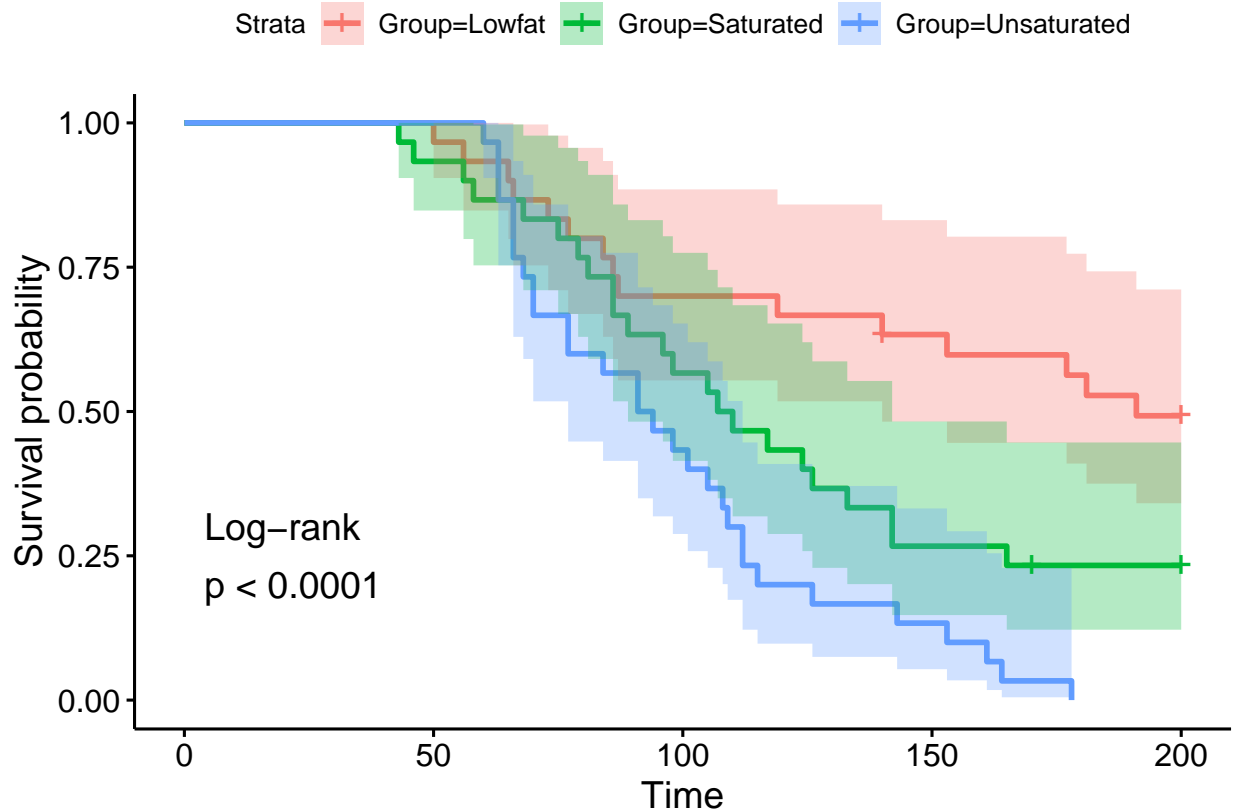
## 1. ¿Qué tipo de censura presentan los datos?

Hay dos tipos de censura:

- por un lado, para algunas ratas el tiempo hasta el suceso del evento fue más largo que el período de seguimiento de 200 días, por lo que no presentaron el evento durante el seguimiento. Esto es una censura por derecha;
- por otro lado, dos ratas murieron accidentalmente antes de desarrollar del tumor y antes de finalizar el período de seguimiento.

2. Graficar las curvas de supervivencia para los tres grupos, estimadas por Kaplan Meier  
¿Cuál es el estimador de la mediana del tiempo de supervivencia y su IC al 95% para cada uno de los grupos?

```
#modelo
fit<-survfit(Surv(time,Censored) ~ Group ,data = dieta)
#plot
ggsurvplot(fit, data=dieta,pval=TRUE,conf.int = TRUE,pval.method = TRUE )
```



Podemos ver que las curvas de sobrevivencia de los tres grupos no son iguales para todos los tiempos tanto gráficamente como con el test Log-rank. En el gráfico podemos ver que para un tiempo un poco menor a los 100 días las curvas de sobrevivencia comienzan a separarse. Por otro lado, el p-valor del Log-rank test dio mucho menor a 0.05, por lo que se rechaza la hipótesis nula que dice que las curvas son iguales para todo t.

```
fit
```

```
## Call: survfit(formula = Surv(time, Censored) ~ Group, data = dieta)
##
##              n events median 0.95LCL 0.95UCL
## Group=Lowfat   30     15  191.0    140     NA
## Group=Saturated 30     23  108.5     89    142
## Group=Unsaturated 30     30   92.5     77    112
```

- Para el grupo “Lowfat” el estimador de la mediana de la sobrevivencia es de 191 días, con un intervalo de confianza del 95% que va de 140 días hasta infinito. El hecho de que el intervalo de confianza vaya hasta infinito tiene que ver con los datos censurados.

```
dieta$Censored <- as.factor(dieta$Censored)
summary(dieta$Censored[dieta$Group == 'Lowfat'])
```

```
## 0 1
## 15 15
```

Podemos ver que la mitad de los datos del grupo de Lowfat son censurados. De estos 15 datos censurados, 14 no llegaron a desarrollar el tumor en los 200 días de seguimiento y 1 murió accidentalmente.

- En el grupo “Saturated” el estimador de la mediana de la sobrevida es de 108.5 días, con un intervalo de confianza del 95% entre 89 días y 142 días.
- El grupo “Unsaturated” tiene una mediana de sobrevida de 92.5 días, con un intervalo de confianza del 95% entre 77 días y 112 días.

### 3. ¿Cuánto vale la probabilidad estimada de supervivencia a 100 días para cada uno de los grupos? Mostrar los IC al 95%

```
summary(fit, times=c(100))
```

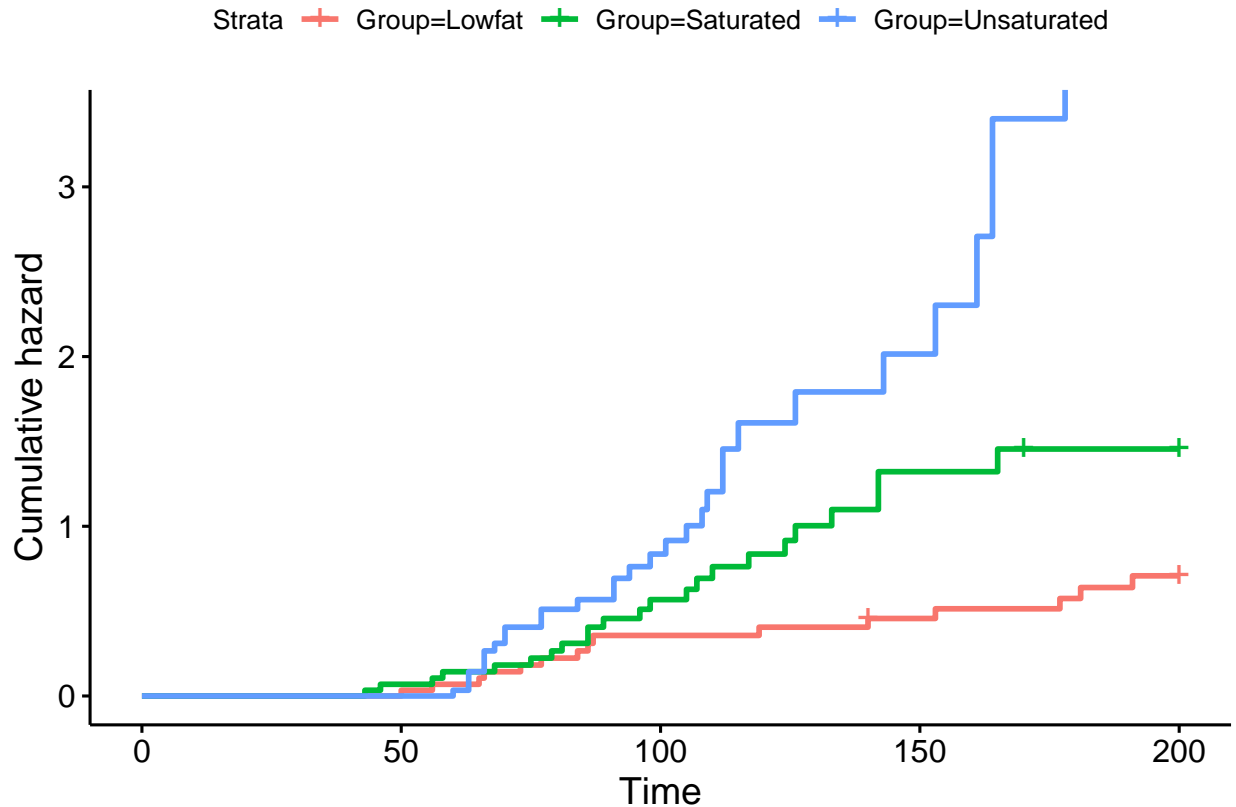
```
## Call: survfit(formula = Surv(time, Censored) ~ Group, data = dieta)
##
##           Group=Lowfat
##      time      n.risk  n.event  survival  std.err lower 95% CI
##    100.0000     21.0000     9.0000     0.7000     0.0837     0.5538
## upper 95% CI
##      0.8848
##
##           Group=Saturated
##      time      n.risk  n.event  survival  std.err lower 95% CI
##    100.0000     17.0000    13.0000     0.5667     0.0905     0.4144
## upper 95% CI
##      0.7749
##
##           Group=Unsaturated
##      time      n.risk  n.event  survival  std.err lower 95% CI
##    100.0000     13.0000    17.0000     0.4333     0.0905     0.2878
## upper 95% CI
##      0.6524
```

- La probabilidad estimada de supervivencia a 100 días para el grupo “Lowfat” es 0.7, con un intervalo de confianza del 95% entre 0.5538 y 0.8848.
- La probabilidad estimada de supervivencia a 100 días para el grupo “Saturated” es 0.5667, con un intervalo de confianza del 95% entre 0.4144 y 0.7749.
- La probabilidad estimada de supervivencia a 100 días para el grupo “Unsaturated” es 0.4333, con un intervalo de confianza del 95% entre 0.2878 y 0.6524.

Podemos ver que si bien la probabilidad de supervivencia es diferente para cada grupo, los tres intervalos de confianza se solapan. Por lo tanto, para el día 100 la probabilidad de supervivencia no es significativamente diferente entre los tres grupos.

4. ¿Los coeficientes entre las funciones de riesgo de los distintos pares de grupos son aproximadamente constantes a través del tiempo?

```
ggsurvplot(fit, data = dieta, fun = "cumhaz")
```



Se puede decir que los coeficientes entre las funciones de riesgo tomando los distintos grupos de a pares se mantienen más o menos constantes a lo largo del tiempo ya que vemos que el compartamiento de las tres curvas es más o menos similar. Es decir, que las tres curvas tienen aproximadamente la misma la tasa de crecimiento, más allá de que el grupo “Unsaturated” presente el mayor riesgo y que el grupo “Lowfat” presente el menor riesgo.

5. Realice el modelo de Cox multivariado y analice sus coeficientes. ¿Qué conclusiones puede sacar?

```
dieta$Censored <- as.numeric(dieta$Censored)
cox<-coxph(Surv(time,Censored) ~ Group ,data = dieta)
summary(cox)
```

```
## Call:
## coxph(formula = Surv(time, Censored) ~ Group, data = dieta)
##
##   n= 90, number of events= 68
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## GroupSaturated  0.7854    2.1932  0.3353  2.342  0.0192 *
```

```
## GroupUnsaturated 1.4287    4.1731    0.3317 4.307 1.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## GroupSaturated      2.193      0.4559      1.137      4.232
## GroupUnsaturated     4.173      0.2396      2.178      7.995
##
## Concordance= 0.624  (se = 0.035 )
## Likelihood ratio test= 20.25  on 2 df,   p=4e-05
## Wald test              = 18.85  on 2 df,   p=8e-05
## Score (logrank) test = 20.87  on 2 df,   p=3e-05
```

El test realizado toma como estado basal, o de comparación, al grupo “Lowfat”, por lo que el modelo teórico es:

$$h(t) = h_0(t)e^{\beta_{saturated}x_{saturated}}e^{\beta_{unsaturated}x_{unsaturated}},$$

donde  $e^{\beta_{saturated}}$  es el hazard ratio de “Saturated” vs “Lowfat” y  $e^{\beta_{unsaturated}}$  es el hazard ratio de “Unsaturated” vs “Lowfat”. Además,  $h_0$  representa el riesgo basal.

- Mantener una dieta “saturated” aumenta 2.193 veces más el riesgo de desarrollar un tumor que si se mantiene una dieta “lowfat”, con un intervalo de confianza del 95% entre 1.137 y 4.232. Esta asociación resulta significativa, con un p-valor menor a 0.05.
- Mantener una dieta “unsaturated” aumenta en un 317.3% el riesgo de desarrollar un tumor si se comapara con una dieta “lowfat”, con un intervalo de confianza del 95% entre 117.8% y 699.5% Esta asociación resulta significativa, con un p-valor menor a 0.05.

Por lo tanto, las dietas “Saturated” y “Unsaturated” son factores de riesgo para el desarrollo de tumor.