



Tarea 6

Modelo de Riesgos Proporcionales de Cox

Modelos de series de Tiempo y Supervivencia

Profesor: Naranjo Albarrán Lizbeth

Adjuntos: Reyes González Belén

Rivas Godoy Yadira

Integrantes: Cuéllar Chávez Eduardo de Jesús

García Tapia Jesús Eduardo

Miranda Meraz Areli Gissell

Ramírez Maciel José Antonio

Saldaña Morales Ricardo

Grupo: 9249

Fecha: 17/DIC/2021

Utiliza la base de datos de R llamada larynx, del paquete KMsurv, y realice lo siguiente:

Ajuste de modelo

Ajuste un modelo de riesgos proporcionales de Cox para definir la contribución de las variables al tiempo de supervivencia de los pacientes.

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

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

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

Hagamos un modelo de “Uno en uno” con el fin de ver cuales son significativos y, posteriormente, hacer el modelo conjunto

Modelo para etapa

Sin modificar

```
#Creamos el objeto
Surv_general <- Surv(larynx$time, larynx$delta, type = "right")

#Modelo de etapas
stage_model <- coxph(Surv_general ~ factor(stage))
stage_model
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage))
##
##              coef exp(coef) se(coef)      z      p
## factor(stage)2 0.06481    1.06696  0.45843 0.141  0.8876
## factor(stage)3 0.61481    1.84930  0.35519 1.731  0.0835
## factor(stage)4 1.73490    5.66838  0.41939 4.137 3.52e-05
##
## Likelihood ratio test=16.49 on 3 df, p=0.0009016
## n= 90, number of events= 50
```

```
#Notamos que la base de comparación es la etapa 1
```

¡Tiene un p-value muy pequeño **conjuntamente**! Ello quiere decir que el coeficiente es significativo, y concuerda con lo obtenido en la tarea 5, donde vimos que para cada una de las etapas, el test de Log-Rank y de Wilcoxon generalizada nos dice que son funciones de supervivencia distintos, por ende es natural que haciendo un modelo de Cox, se tenga un parámetro significativo; sin embargo esto es de **manera conjunta**. Ya que el p-value del grupo de la etapa 2 es demasiado grande, probemos con las nuevas clasificaciones que probamos en la tarea 5:

Agrupando la etapa 1 y 2:

```
stage_aux <- larynx %>%
  mutate(stage_aux = factor(ifelse(stage %in% c(1,2), "1 o 2", stage)))
stage_fit_aux <- survfit(Surv_general ~ stage_aux, type = "kaplan-meier",
                        conf.int = 0.95, conf.type = "plain", data = stage_aux)

stage_model_aux1 <- coxph(Surv_general ~ factor(stage_aux), data = stage_aux)
stage_model_aux1
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux), data = stage_aux)
##
##               coef exp(coef) se(coef)      z      p
## factor(stage_aux)3 0.5946    1.8124   0.3238 1.836 0.0663
## factor(stage_aux)4 1.7141    5.5516   0.3915 4.378 1.2e-05
##
## Likelihood ratio test=16.47 on 2 df, p=0.0002658
## n= 90, number of events= 50
#la etapa base es la 1 o 2
```

¡De manera conjunta tiene un p-value muy pequeño! y notemos que, si bien en la etapa 3 tiene un p-value de 0.0663, este es menor a 0.2, y vimos con la profesora que, acorde a autores como Hosmer-Lemeshow, si tienen un p-value menor o igual a 0.2, se deben considerar en el modelo múltiple

Agrupando la etapa 1 y 2 en un grupo, y la 3 y 4 en otro:

```
stage_aux2<- larynx %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"Temprana(1,2)",
                                "Avanzada(3,4)")))
stage_fit_aux <- survfit(Surv_general ~ stage_aux, type = "kaplan-meier",
                        conf.int = 0.95, conf.type = "plain",data=stage_aux2)
stage_model_aux2<-coxph(Surv_general~factor(stage_aux),data=stage_aux2)
stage_model_aux2
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux), data = stage_aux2)
##
##               coef exp(coef) se(coef)      z      p
## factor(stage_aux)Temprana(1,2) -0.8846    0.4129   0.2869 -3.084 0.00204
##
## Likelihood ratio test=9.54 on 1 df, p=0.002009
## n= 90, number of events= 50
#La etapa base es la avanzada
```

Teneos un p-value demasiado pequeño, por lo que definitivamente es una variable significativa bajo esta clasificación

Modelo para edad

Sin modificar (continúa)

```
age_model<-coxph(Surv_general~age)
age_model
```

```
## Call:
## coxph(formula = Surv_general ~ age)
##
##           coef exp(coef) se(coef)      z      p
## age 0.02328    1.02356   0.01449 1.607 0.108
##
## Likelihood ratio test=2.63 on 1 df, p=0.1048
## n= 90, number of events= 50
```

Tiene un p-value mayor a 0.05, sin embargo, en clase vimos que autores como Hosmer-Lemeshow consideran que, en un análisis exploratorio, las variables con un nivel de significancia menor o igual a 0.2 se consideren en el modelo múltiple, entonces podemos ingresar la variable de edad al modelo múltiple; ello coincide un poco con lo que obtuvimos en la tarea 5: Que dada cierta clasificación en el grupo de edad, sí había una diferencia significativa con base en los test.

Agrupando de 40 a 70 y de 71 en adelante

Por curiosidad, veamos qué pasa con la clasificación que propusimos en la tarea 5:

```
aux_age<- larynx %>%
  mutate("Grupo"=cut(larynx$age,breaks=c(40,70,Inf)))
aux_age_model<- coxph(Surv_general~aux_age$Grupo,data = aux_age)
aux_age_model
```

```
## Call:
## coxph(formula = Surv_general ~ aux_age$Grupo, data = aux_age)
##
##               coef exp(coef) se(coef)      z      p
## aux_age$Grupo(70,Inf] 0.6125    1.8451   0.2893  2.117 0.0342
##
## Likelihood ratio test=4.25 on 1 df, p=0.0393
## n= 90, number of events= 50
```

#Notamos que la base es el grupo de 40 a 70

Podemos observar que bajo la clasificación en la que obtuvimos una función de supervivencia distinta para ambos grupos en la tarea 5, tenemos aquí un p-value menor a 0.05, por lo que concuerda con lo obtenido en los test de la tarea anterior.

```
diagyr_model<-coxph(Surv_general~diagyr)
diagyr_model
```

```
## Call:
## coxph(formula = Surv_general ~ diagyr)
##
##               coef exp(coef) se(coef)      z      p
## diagyr 0.02200    1.02225   0.07173  0.307 0.759
##
## Likelihood ratio test=0.09 on 1 df, p=0.7588
## n= 90, number of events= 50
```

Definitivamente, no es una variable que aporte mucho: Su p-value es altísimo, por lo que los coeficientes no son significativos inclusive bajo el criterio del 0.2 propuesto por algunos autores como Hosmer-Lemeshow, por lo que no la consideraremos en el modelo conjunto.

Modelo conjunto

Veamos el modelo en conjunto:

Modelo 1: Sin modificaciones

```
modelo_conjunto<-coxph(Surv_general ~ factor(stage)+age)
modelo_conjunto
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage) + age)
```

```
##
##               coef exp(coef) se(coef)      z      p
## factor(stage)2 0.14004    1.15032  0.46249 0.303   0.7620
## factor(stage)3 0.64238    1.90100  0.35611 1.804   0.0712
## factor(stage)4 1.70598    5.50678  0.42191 4.043 5.27e-05
## age            0.01903    1.01921  0.01426 1.335   0.1820
##
## Likelihood ratio test=18.31 on 4 df, p=0.001072
## n= 90, number of events= 50
```

¡Conjuntamente, son significativos! El p-value es menor a 0.05; sin embargo la edad y 2 de las etapas de la enfermedad tienen un p.value mayor a 0.05, por lo que no son significativos de manera individual.

Modelo 2: Con grupo de edad de 40 a 70 y de 71 en adelante

Y usando la clasificación de edades propuesta, hubiésemos tenido este modelo:

```
modelo_conjunto_aux<-coxph(Surv_general ~ factor(stage)+Grupo,data=aux_age)
modelo_conjunto_aux
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage) + Grupo, data = aux_age)
##
##               coef exp(coef) se(coef)      z      p
## factor(stage)2 0.1406    1.1510  0.4611 0.305   0.7604
## factor(stage)3 0.7118    2.0378  0.3605 1.974   0.0483
## factor(stage)4 1.6544    5.2297  0.4225 3.916 9.01e-05
## Grupo(70,Inf] 0.5497    1.7328  0.3015 1.823   0.0683
##
## Likelihood ratio test=19.68 on 4 df, p=0.0005781
## n= 90, number of events= 50
```

De la misma manera, y como se esperaba, usando los grupos propuestos para edad, los coeficientes son significativos de manera conjunta. De hecho, el p-value es menor que en el primer modelo conjunto obtenido en la línea anterior. Sin embargo, tenemos parámetros no significativos en las etapas de la enfermedad

Modelo 3: Con grupo de edad de 40 a 70 y de 71 en adelante; así como etapa 1 y 2 agrupadas

```
stage_aux<- aux_age %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"1 o 2",stage)))
modelo_conjunto_aux1<-coxph(Surv_general ~ stage_aux+Grupo,data=stage_aux)
modelo_conjunto_aux1
```

```
## Call:
## coxph(formula = Surv_general ~ stage_aux + Grupo, data = stage_aux)
##
##               coef exp(coef) se(coef)      z      p
## stage_aux3      0.6678    1.9499  0.3274 2.040   0.0414
## stage_aux4      1.6112    5.0088  0.3956 4.073 4.65e-05
## Grupo(70,Inf] 0.5408    1.7175  0.3000 1.803   0.0714
##
## Likelihood ratio test=19.59 on 3 df, p=0.0002067
## n= 90, number of events= 50
```

Con este modelo son significativos de manera conjunta, sin embargo en la edad tenemos un coeficiente no significativo al 95% de confianza, pero al 90% sí lo sería.

Modelo 4: Con grupo de edad de 40 a 70 y de 71 en adelante; etapas agrupadas en temprana y avanzada

```
stage_aux_2<- aux_age %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"Temprana(1,2)",
                                "Avanzada(3,4)")))
modelo_conjunto_aux2<-coxph(Surv_general ~ stage_aux+Grupo,data=stage_aux_2)
modelo_conjunto_aux2
```

```
## Call:
## coxph(formula = Surv_general ~ stage_aux + Grupo, data = stage_aux_2)
##
##               coef exp(coef) se(coef)      z      p
## stage_auxTemprana(1,2) -0.9381    0.3914  0.2889 -3.247 0.00117
## Grupo(70,Inf]          0.6895    1.9928  0.2911  2.369 0.01785
##
## Likelihood ratio test=14.82 on 2 df, p=0.0006041
## n= 90, number of events= 50
```

En este modelo, todos los parámetros son significativos de manera individual y de manera conjunta (Bajo el grupo de edad de 40 a 70 y de 71 en adelante, y con etapa temprana y avanzada)

Ejercicio 1 : Análisis de los coeficientes de regresión

¿Cuál es la estimación puntual para los coeficientes de regresión? Interprete los coeficientes de regresión.

Estimación puntual e interpretación

Antes de entrar a este apartado, ya calculamos los modelos, retomémoslos:

Modelo 1: Sin modificaciones

```
modelo_conjunto$coefficients

## factor(stage)2 factor(stage)3 factor(stage)4      age
##      0.1400402      0.6423817      1.7059796      0.0190311
```

Modelo 2: Con grupo de edad de 40 a 70 y de 71 en adelante

```
modelo_conjunto_aux$coefficients

## factor(stage)2 factor(stage)3 factor(stage)4 Grupo(70,Inf]
##      0.1406373      0.7118470      1.6543550      0.5497304
```

Modelo 3: Con grupo de edad de 40 a 70 y de 71 en adelante; así como etapa 1 y 2 agrupadas

```
modelo_conjunto_aux1$coefficients

##      stage_aux3      stage_aux4 Grupo(70,Inf]
##      0.6677627      1.6111884      0.5408500
```

Modelo 4: Con grupo de edad de 40 a 70 y de 71 en adelante; etapas agrupadas en temprana y avanzada

```
modelo_conjunto_aux2$coefficients

## stage_auxTemprana(1,2)      Grupo(70,Inf]
##                -0.9381336          0.6895187
```

Ejercicio 2: Análisis del efecto de las variables en el modelo

¿Las variables explicativas tienen o no efecto en el modelo? Obtenga un intervalo de confianza al 95% para la estimación de los coeficientes de regresión. Justifique.

Intervalo del 95% de confianza y Efecto/significancia de las variables explicativas en el modelo

Modelo 1: Sin modificaciones

```
confint(modelo_conjunto)

##                2.5 %      97.5 %
## factor(stage)2 -0.766415942  1.04649625
## factor(stage)3 -0.055582140  1.34034559
## factor(stage)4  0.879044677  2.53291454
## age            -0.008914884  0.04697709
```

Podemos observar que la etapa 2 y 3, así como la edad (En este modelo), contienen al cero en su intervalo de confianza al 95%, por lo que bajo este contexto no son significativas. Ello puede ser debido a la clasificación que estamos haciendo, ya que como vimos en la tarea 5, se podían juntar las etapas 1 y 2, además de que convenía juntar la edad de 40 a 70 y de 71 en adelante. El único coeficiente que aporta algo en este modelo es, entonces, la etapa 4; notamos que la etapa 3 está casi del lado positivo únicamente, por lo que a un nivel de confianza menor sí sería significativo, al igual que la edad. En resumen: En este contexto, convendría tomar la edad como grupo y no continua, y ver si de esta manera ya tenemos un parámetro significativo, que aporte al modelo. En cuanto a las etapas, definitivamente la etapa 2 no aporta mucho al modelo, tomando como base la etapa 1, esto por lo que vimos que prácticamente no había mucha diferencia en la función de supervivencia. La etapa 4 sí difiere mucho, esto tiene sentido puesto que es la etapa más avanzada, por ende es natural que tenga un comportamiento distinto un paciente en etapa avanzada a comparación de uno en etapa temprana.

Modelo 2: Con grupo de edad de 40 a 70 y de 71 en adelante

```
confint(modelo_conjunto_aux)

##                2.5 %      97.5 %
## factor(stage)2 -0.763188207  1.044463
## factor(stage)3  0.005194658  1.418499
## factor(stage)4  0.826298180  2.482412
## Grupo(70,Inf] -0.041230381  1.140691
```

Aquí la etapa 4 y 3 tienen ya un coeficiente significativo, por lo que al agrupar las edades el parámetro de la etapa 3 pasó a ser significativo, sin embargo seguimos teniendo problemas con la etapa 2, y vemos que el coeficiente de la edad por poco podría ser significativo, ya que está más del lado positivo que del negativo, por mucho. Entonces podríamos tratar de agrupar la etapa 1 y 2 para ver qué efecto tiene

Modelo 3: Con grupo de edad de 40 a 70 y de 71 en adelante; así como etapa 1 y 2 agrupadas

```
confint(modelo_conjunto_aux1)
```

```
##                2.5 %    97.5 %  
## stage_aux3      0.02614535 1.309380  
## stage_aux4      0.83582143 2.386555  
## Grupo(70,Inf] -0.04706828 1.128768
```

Agrupando las etapas 1 y 2, así como la edad, vemos que seguimos teniendo a la etapa 3 y 4 como parámetros significativos. ¿Tendrá algún efecto el considerar los grupos 3 y 4 en conjunto?

Modelo 4: Con grupo de edad de 40 a 70 y de 71 en adelante; etapas agrupadas en temprana y avanzada

```
confint(modelo_conjunto_aux2)
```

```
##                2.5 %    97.5 %  
## stage_auxTemprana(1,2) -1.5044019 -0.3718654  
## Grupo(70,Inf]          0.1189824  1.2600550
```

De esta manera, vemos que ya los dos parámetros son significativos, al agrupar la etapa 1 y 2 en un conjunto, y la 3 y 4 en otro; así como la edad. Esto tiene sentido ya que, el primer grupo de edad es, relativamente “joven” o, bueno, aún son adultos que podrían tener la fuerza para luchar con la enfermedad, respecto al segundo grupo que son personas ya en etapa de vejez. Y las etapas 1 y 2 son las iniciales, cuando las 3 y 4 es estar ya en una etapa algo avanzada de la enfermedad, por ello es natural que sí sean significativos los parámetros.

Ejercicio 3: Análisis de supuesto de riesgos proporcionales

¿Es válido tu modelo de acuerdo al supuesto de riesgos proporcionales? Realice las pruebas de análisis de residuos y concluya.

Pruebas de análisis de residuos

Modelo 1: Sin modificaciones

Modelo 2: Con grupo de edad de 40 a 70 y de 71 en adelante

Modelo 3: Con grupo de edad de 40 a 70 y de 71 en adelante; así como etapa 1 y 2 agrupadas

Modelo 4: Con grupo de edad de 40 a 70 y de 71 en adelante; etapas agrupadas en temprana y avanzada