



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
## Warning: package 'muhaz' was built under R version 4.1.2
## Warning: package 'TH.data' was built under R version 4.1.2
## Warning: package 'ggfortify' was built under R version 4.1.2
## Warning: package 'proto' 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))
summary(stage_model)

## Call:
## coxph(formula = Surv_general ~ factor(stage))
##
##      n= 90, number of events= 50
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## 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 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(stage)2    1.067    0.9372    0.4344    2.62
## factor(stage)3    1.849    0.5407    0.9219    3.71
## factor(stage)4    5.668    0.1764    2.4916   12.90
##
## Concordance= 0.668 (se = 0.037 )
## Likelihood ratio test= 16.49 on 3 df,  p=9e-04
## Wald test               = 19.24 on 3 df,  p=2e-04
## Score (logrank) test = 22.88 on 3 df,  p=4e-05
```

#Notamos que la base de comparación es la etapa 1

No tenemos problema con el coeficiente de la etapa 4, y bajo lo que se vio en clase, que acorde a algunos autores en supervivencia 0.2 sigue siendo significativo, tampoco lo hay para la etapa. Sin embargo, ya que el p-value del grupo de la etapa 2 es demasiado grande, por lo que descartamos bajo esta clasificación. Ahora 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)
summary(stage_model_aux1)
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux), data = stage_aux)
##
##    n= 90, number of events= 50
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## 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 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(stage_aux)3    1.812    0.5518    0.9607    3.419
## factor(stage_aux)4    5.552    0.1801    2.5774   11.958
##
## Concordance= 0.664  (se = 0.036 )
## Likelihood ratio test= 16.47  on 2 df,   p=3e-04
## Wald test               = 19.23  on 2 df,   p=7e-05
## Score (logrank) test = 22.86  on 2 df,   p=1e-05
```

#la etapa base es la 1 o 2

¡Resolvimos el problema! 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)
summary(stage_model_aux2)
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux), data = stage_aux2)
```

```
##
##   n= 90, number of events= 50
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(stage_aux)Temprana(1,2) -0.8846    0.4129    0.2869 -3.084  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(stage_aux)Temprana(1,2)    0.4129      2.422    0.2353    0.7244
##
## Concordance= 0.641 (se = 0.035 )
## Likelihood ratio test= 9.54  on 1 df,  p=0.002
## Wald test              = 9.51  on 1 df,  p=0.002
## Score (logrank) test = 10.12 on 1 df,  p=0.001
```

#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)
summary(age_model)

## Call:
## coxph(formula = Surv_general ~ age)
##
##   n= 90, number of events= 50
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## age 0.02328   1.02356   0.01449  1.607   0.108
##
##      exp(coef) exp(-coef) lower .95 upper .95
## age      1.024      0.977   0.9949   1.053
##
## Concordance= 0.555 (se = 0.046 )
## Likelihood ratio test= 2.63  on 1 df,  p=0.1
## Wald test              = 2.58  on 1 df,  p=0.1
## Score (logrank) test = 2.6   on 1 df,  p=0.1
```

Tiene un p-value mayor a 0.05, sin embargo, en clase vimos que autores como Hosmer-Lemeshow consideran que, en un analisis exploratorio, las variables con un nivel de significancia menor o igual a 0.2 se consideren en el modelo multiple, 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)
summary(diagyr_model)
```

```
## Call:
## coxph(formula = Surv_general ~ diagyr)
##
##      n= 90, number of events= 50
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## diagyr 0.02200    1.02225    0.07173 0.307    0.759
##
##      exp(coef) exp(-coef) lower .95 upper .95
## diagyr      1.022      0.9782    0.8882    1.177
##
## Concordance= 0.528 (se = 0.043 )
## Likelihood ratio test= 0.09 on 1 df,  p=0.8
## Wald test               = 0.09 on 1 df,  p=0.8
## Score (logrank) test = 0.09 on 1 df,  p=0.8
```

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: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_1<-coxph(Surv_general ~ factor(stage_aux)+age,data=stage_aux)
summary(modelo_conjunto_1)
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux) + age, data = stage_aux)
##
##      n= 90, number of events= 50
```

```
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(stage_aux)3 0.59964   1.82145  0.32400 1.851   0.0642 .
## factor(stage_aux)4 1.66290   5.27457  0.39469 4.213 2.52e-05 ***
## age                0.01851   1.01868  0.01413 1.310   0.1903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(stage_aux)3     1.821     0.5490   0.9652   3.437
## factor(stage_aux)4     5.275     0.1896   2.4334  11.433
## age                    1.019     0.9817   0.9909   1.047
##
## Concordance= 0.679 (se = 0.039 )
## Likelihood ratio test= 18.22 on 3 df,  p=4e-04
## Wald test              = 21.05 on 3 df,  p=1e-04
## Score (logrank) test = 24.69 on 3 df,  p=2e-05
```

Todos son significativos bajo el criterio de tener un p-value menor o igual a 0.2

Modelo 2: etapas “Avanzada” y “Temprana”

```
stage_aux_2<- aux_age %>%
  mutate(stage_aux=factor(ifelse(stage %in% c(1,2),"Temprana(1,2)",
                                "Avanzada(3,4)")))
modelo_conjunto_2<-coxph(Surv_general ~ factor(stage_aux)+age,data=stage_aux_2)
summary(modelo_conjunto_2)
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux) + age, data = stage_aux_2)
##
##      n= 90, number of events= 50
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## factor(stage_aux)Temprana(1,2) -0.87939  0.41504  0.28701 -3.064  0.00218 **
## age                          0.02286   1.02313  0.01457  1.570  0.11649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## factor(stage_aux)Temprana(1,2)   0.415     2.4094   0.2365   0.7284
## age                            1.023     0.9774   0.9943   1.0528
##
## Concordance= 0.664 (se = 0.039 )
## Likelihood ratio test= 12.05 on 2 df,  p=0.002
## Wald test              = 11.93 on 2 df,  p=0.003
## Score (logrank) test = 12.57 on 2 df,  p=0.002
```

Vemos que aquí mejoran los p-values, pues se hacen más pequeños

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

```
modelo_conjunto_3<-coxph(Surv_general ~ factor(stage_aux)+Grupo,data=stage_aux_2)
modelo_conjunto_3
```

```
## Call:
## coxph(formula = Surv_general ~ factor(stage_aux) + Grupo, data = stage_aux_2)
##
##               coef exp(coef) se(coef)      z      p
## factor(stage_aux)Temprana(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 inclusive a un nivel del 95% de confianza

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:etapa 1 y 2 agrupadas

```
modelo_conjunto_1$coefficients
```

```
## factor(stage_aux)3 factor(stage_aux)4      age
##      0.59963511      1.66289670      0.01850742
```

En cuanto a la edad, esto nos quiere decir que mientras la edad aumenta, presentamos un mayor riesgo de morir al padecer cáncer, es decir, es directamente proporcional.

Lo mismo para las etapas: La etapa 3 tiene un mayor riesgo directamente proporcional de morir al padecer cáncer respecto a la etapa 1 y 2.

La etapa 4 presenta un riesgo aún mayor (respecto al presentado por la etapa 3, ya que este coeficiente es mayor que el de 3) e igualmente directamente proporcional de morir al padecer cáncer, respecto a las etapas 1 y 2.

Modelo 2: etapas “Avanzada” y “Temprana”

```
modelo_conjunto_2$coefficients
```

```
## factor(stage_aux)Temprana(1,2)      age
##      -0.87938583      0.02286372
```

Para la edad tenemos la misma interpretación: Nos quiere decir que mientras la edad aumenta, presentamos un mayor riesgo de morir al padecer cáncer, es decir, es directamente proporcional.

Notemos que aquí el signo en el coeficiente de etapa cambia, pero es porque ahora está respecto a la etapa 3 y 4 agrupadas, por lo que la interpretación sigue teniendo sentido: Hay un riesgo menor en la etapa 1 y 2 respecto a la 3 y 4, ya que se presenta un riesgo inversamente proporcional de morir al padecer cáncer

Modelo 3:etapas 1 y 2 agrupadas

```
modelo_conjunto_3$coefficients
```

```
## factor(stage_aux)Temprana(1,2)      Grupo(70,Inf]
##      -0.9381336      0.6895187
```

Las mismas interpretaciones:

El grupo de 71 años en adelante guarda una relación directamente proporcional en cuanto a riesgo de fallecer por cáncer respecto al grupo de adultos de 40 a 70 años.

La interpretación para las etaoas se mantiene igual: Hay un riesgo menor en la etapa 1 y 2 respecto a la 3 y 4, ya que se presenta un riesgo inversamente proporcional de morir al padecer cáncer

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:etapa 1 y 2 agrupadas

```
confint(modelo_conjunto_1)

##                2.5 %      97.5 %
## factor(stage_aux)3 -0.035385255  1.23465547
## factor(stage_aux)4  0.889309119  2.43648429
## age                -0.009186478  0.04620132
```

Podemos observar que en este caso, a un 95% de confianza la edad y la etapa 3 no parecen ser significativas. Sin embargo, recordemos que esto puede ser debido a que no contamos con tantos datos como debiese ser necesario (Tenemos 40 censuras de 90 datos disponibles), por ello es que pueden pasar este tipo de cosas y se recomienda tener un nivel de confianza del 80% en lugar del 95%, por lo que entendimos en clase.

Igualmente, notamos que el coeficiente de etapa 3 está más pegado su intervalo a la derecha que a la izquierda, lo mismo para la edad.

Como extra: ¿Interactúan?

```
survdifff(Surv_general ~ age + factor(stage_aux), rho = 0,
          data = stage_aux)

## Call:
## survdifff(formula = Surv_general ~ age + factor(stage_aux), data = stage_aux,
##           rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## age=41, factor(stage_aux)=4      1          1  0.1670  4.15e+00  4.26e+00
## age=43, factor(stage_aux)=1 o 2  1          1  0.4438  6.97e-01  7.17e-01
## age=45, factor(stage_aux)=1 o 2  1          1  0.3373  1.30e+00  1.33e+00
## age=47, factor(stage_aux)=1 o 2  2          0  1.2482  1.25e+00  1.31e+00
## age=48, factor(stage_aux)=1 o 2  1          0  0.5927  5.93e-01  6.13e-01
## age=48, factor(stage_aux)=4      1          0  0.5927  5.93e-01  6.13e-01
## age=49, factor(stage_aux)=3      3          2  0.8456  1.58e+00  1.64e+00
## age=50, factor(stage_aux)=1 o 2  1          0  1.0725  1.07e+00  1.12e+00
## age=51, factor(stage_aux)=1 o 2  2          0  0.7589  7.59e-01  7.88e-01
## age=51, factor(stage_aux)=3      1          0  1.1836  1.18e+00  1.25e+00
## age=52, factor(stage_aux)=1 o 2  1          1  0.5695  3.25e-01  3.37e-01
## age=52, factor(stage_aux)=3      1          0  0.4830  4.83e-01  4.97e-01
## age=53, factor(stage_aux)=1 o 2  2          1  1.2659  5.58e-02  5.85e-02
## age=53, factor(stage_aux)=3      1          1  0.2769  1.89e+00  1.93e+00
```



```

## age=54, factor(stage_aux)=3      2      1  0.9634  1.39e-03  1.45e-03
## age=56, factor(stage_aux)=1 o 2 1      0  1.1836  1.18e+00  1.25e+00
## age=57, factor(stage_aux)=1 o 2 1      0  0.3373  3.37e-01  3.46e-01
## age=57, factor(stage_aux)=3      1      1  0.0801  1.06e+01  1.08e+01
## age=58, factor(stage_aux)=1 o 2 3      1  2.2098  6.62e-01  7.05e-01
## age=59, factor(stage_aux)=3      1      1  0.6221  2.29e-01  2.37e-01
## age=60, factor(stage_aux)=1 o 2 1      1  0.4438  6.97e-01  7.17e-01
## age=60, factor(stage_aux)=3      1      1  0.1933  3.37e+00  3.45e+00
## age=61, factor(stage_aux)=1 o 2 2      0  2.1125  2.11e+00  2.27e+00
## age=62, factor(stage_aux)=1 o 2 1      1  0.9956  1.95e-05  2.04e-05
## age=62, factor(stage_aux)=4      1      1  0.3217  1.43e+00  1.47e+00
## age=63, factor(stage_aux)=1 o 2 3      2  1.2640  4.29e-01  4.51e-01
## age=63, factor(stage_aux)=3      1      0  0.5927  5.93e-01  6.13e-01
## age=64, factor(stage_aux)=1 o 2 3      1  1.5767  2.11e-01  2.22e-01
## age=64, factor(stage_aux)=3      1      1  0.2206  2.75e+00  2.82e+00
## age=65, factor(stage_aux)=3      2      1  1.8052  3.59e-01  3.84e-01
## age=65, factor(stage_aux)=4      2      2  0.1525  2.24e+01  2.29e+01
## age=66, factor(stage_aux)=1 o 2 2      0  1.6177  1.62e+00  1.71e+00
## age=66, factor(stage_aux)=3      1      0  0.5927  5.93e-01  6.13e-01
## age=67, factor(stage_aux)=1 o 2 2      1  1.2663  5.60e-02  5.85e-02
## age=68, factor(stage_aux)=1 o 2 3      1  2.8489  1.20e+00  1.31e+00
## age=68, factor(stage_aux)=3      1      1  1.1836  2.85e-02  3.00e-02
## age=68, factor(stage_aux)=4      1      1  0.2069  3.04e+00  3.11e+00
## age=69, factor(stage_aux)=3      2      0  1.8058  1.81e+00  1.91e+00
## age=69, factor(stage_aux)=4      1      1  0.3063  1.57e+00  1.61e+00
## age=70, factor(stage_aux)=1 o 2 2      1  1.1384  1.68e-02  1.76e-02
## age=70, factor(stage_aux)=3      1      1  0.7811  6.14e-02  6.35e-02
## age=71, factor(stage_aux)=1 o 2 1      0  0.3063  3.06e-01  3.14e-01
## age=71, factor(stage_aux)=3      1      1  0.0564  1.58e+01  1.62e+01
## age=71, factor(stage_aux)=4      2      2  0.5394  3.95e+00  4.08e+00
## age=72, factor(stage_aux)=1 o 2 1      0  0.4830  4.83e-01  4.97e-01
## age=72, factor(stage_aux)=3      1      1  0.2769  1.89e+00  1.93e+00
## age=73, factor(stage_aux)=1 o 2 2      0  2.2561  2.26e+00  2.45e+00
## age=74, factor(stage_aux)=1 o 2 1      1  0.7356  9.50e-02  9.83e-02
## age=74, factor(stage_aux)=3      1      1  0.2483  2.28e+00  2.33e+00
## age=74, factor(stage_aux)=4      1      0  0.3373  3.37e-01  3.46e-01
## age=75, factor(stage_aux)=1 o 2 1      1  0.6939  1.35e-01  1.40e-01
## age=76, factor(stage_aux)=1 o 2 1      1  0.3882  9.64e-01  9.88e-01
## age=76, factor(stage_aux)=4      1      1  0.0682  1.27e+01  1.30e+01
## age=77, factor(stage_aux)=1 o 2 3      2  1.6624  6.86e-02  7.25e-02
## age=78, factor(stage_aux)=3      1      0  1.1836  1.18e+00  1.25e+00
## age=78, factor(stage_aux)=4      1      1  0.1414  5.21e+00  5.34e+00
## age=79, factor(stage_aux)=1 o 2 1      0  0.9289  9.29e-01  9.69e-01
## age=79, factor(stage_aux)=3      1      1  0.1044  7.69e+00  7.84e+00
## age=81, factor(stage_aux)=1 o 2 2      2  1.2249  4.90e-01  5.15e-01
## age=81, factor(stage_aux)=3      1      1  0.4438  6.97e-01  7.17e-01
## age=82, factor(stage_aux)=3      1      1  0.1414  5.21e+00  5.34e+00
## age=84, factor(stage_aux)=4      1      1  0.5043  4.87e-01  5.02e-01
## age=86, factor(stage_aux)=1 o 2 2      2  0.6151  3.12e+00  3.22e+00
##
## Chisq= 161 on 62 degrees of freedom, p= 1e-10
survdifff(Surv_general ~ age + factor(stage_aux), rho = 1,
          data = stage_aux)

```

```
## Call:
## survdiff(formula = Surv_general ~ age + factor(stage_aux), data = stage_aux,
##          rho = 1)
##
##
```

		N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
## age=41, factor(stage_aux)=4		1	0.867	0.1556	3.25e+00	3.57e+00
## age=43, factor(stage_aux)=1	o 2 1	0.675	0.3625	2.69e-01	3.34e-01	
## age=45, factor(stage_aux)=1	o 2 1	0.722	0.2892	6.48e-01	7.66e-01	
## age=47, factor(stage_aux)=1	o 2 2	0.000	0.9391	9.39e-01	1.27e+00	
## age=48, factor(stage_aux)=1	o 2 1	0.000	0.4526	4.53e-01	5.94e-01	
## age=48, factor(stage_aux)=4		1	0.000	0.4526	4.53e-01	5.94e-01
## age=49, factor(stage_aux)=3		3	1.844	0.6799	1.99e+00	2.50e+00
## age=50, factor(stage_aux)=1	o 2 1	0.000	0.6664	6.66e-01	1.02e+00	
## age=51, factor(stage_aux)=1	o 2 2	0.000	0.6379	6.38e-01	7.79e-01	
## age=51, factor(stage_aux)=3		1	0.000	0.7035	7.03e-01	1.11e+00
## age=52, factor(stage_aux)=1	o 2 1	0.599	0.4396	5.81e-02	7.57e-02	
## age=52, factor(stage_aux)=3		1	0.000	0.3875	3.87e-01	4.87e-01
## age=53, factor(stage_aux)=1	o 2 2	0.844	0.8442	6.84e-08	9.75e-08	
## age=53, factor(stage_aux)=3		1	0.778	0.2444	1.16e+00	1.34e+00
## age=54, factor(stage_aux)=3		2	0.711	0.7656	3.92e-03	5.01e-03
## age=56, factor(stage_aux)=1	o 2 1	0.000	0.7035	7.03e-01	1.11e+00	
## age=57, factor(stage_aux)=1	o 2 1	0.000	0.2892	2.89e-01	3.42e-01	
## age=57, factor(stage_aux)=3		1	0.933	0.0778	9.41e+00	9.90e+00
## age=58, factor(stage_aux)=1	o 2 3	0.711	1.5029	4.17e-01	6.00e-01	
## age=59, factor(stage_aux)=3		1	0.547	0.4687	1.32e-02	1.75e-02
## age=60, factor(stage_aux)=1	o 2 1	0.675	0.3625	2.69e-01	3.34e-01	
## age=60, factor(stage_aux)=3		1	0.844	0.1778	2.50e+00	2.78e+00
## age=61, factor(stage_aux)=1	o 2 2	0.000	1.3163	1.32e+00	2.04e+00	
## age=62, factor(stage_aux)=1	o 2 1	0.387	0.6386	9.90e-02	1.48e-01	
## age=62, factor(stage_aux)=4		1	0.733	0.2779	7.46e-01	8.77e-01
## age=63, factor(stage_aux)=1	o 2 3	1.355	1.0313	1.02e-01	1.28e-01	
## age=63, factor(stage_aux)=3		1	0.000	0.4526	4.53e-01	5.94e-01
## age=64, factor(stage_aux)=1	o 2 3	0.800	1.2016	1.34e-01	1.79e-01	
## age=64, factor(stage_aux)=3		1	0.811	0.2000	1.87e+00	2.10e+00
## age=65, factor(stage_aux)=3		2	0.452	1.2041	4.70e-01	6.98e-01
## age=65, factor(stage_aux)=4		2	1.900	0.1444	2.13e+01	2.31e+01
## age=66, factor(stage_aux)=1	o 2 2	0.000	1.1074	1.11e+00	1.59e+00	
## age=66, factor(stage_aux)=3		1	0.000	0.4526	4.53e-01	5.94e-01
## age=67, factor(stage_aux)=1	o 2 2	0.409	0.9020	2.70e-01	3.71e-01	
## age=68, factor(stage_aux)=1	o 2 3	0.361	1.8226	1.17e+00	1.80e+00	
## age=68, factor(stage_aux)=3		1	0.334	0.7035	1.95e-01	3.06e-01
## age=68, factor(stage_aux)=4		1	0.822	0.1889	2.12e+00	2.37e+00
## age=69, factor(stage_aux)=3		2	0.000	1.1722	1.17e+00	1.74e+00
## age=69, factor(stage_aux)=4		1	0.756	0.2667	8.96e-01	1.05e+00
## age=70, factor(stage_aux)=1	o 2 2	0.638	0.8739	6.40e-02	8.44e-02	
## age=70, factor(stage_aux)=3		1	0.473	0.5483	1.03e-02	1.44e-02
## age=71, factor(stage_aux)=1	o 2 1	0.000	0.2667	2.67e-01	3.12e-01	
## age=71, factor(stage_aux)=3		1	0.978	0.0556	1.53e+01	1.59e+01
## age=71, factor(stage_aux)=4		2	1.615	0.4430	3.10e+00	3.82e+00
## age=72, factor(stage_aux)=1	o 2 1	0.000	0.3875	3.87e-01	4.87e-01	
## age=72, factor(stage_aux)=3		1	0.778	0.2444	1.16e+00	1.34e+00
## age=73, factor(stage_aux)=1	o 2 2	0.000	1.3699	1.37e+00	2.17e+00	
## age=74, factor(stage_aux)=1	o 2 1	0.494	0.5268	2.06e-03	2.85e-03	
## age=74, factor(stage_aux)=3		1	0.800	0.2222	1.50e+00	1.71e+00

```
## age=74, factor(stage_aux)=4      1      0.000  0.2892  2.89e-01  3.42e-01
## age=75, factor(stage_aux)=1 o 2 1      0.514  0.5062  1.07e-04  1.45e-04
## age=76, factor(stage_aux)=1 o 2 1      0.687  0.3250  4.03e-01  4.88e-01
## age=76, factor(stage_aux)=4      1      0.944  0.0667  1.16e+01  1.21e+01
## age=77, factor(stage_aux)=1 o 2 3      1.374  1.1864  2.96e-02  4.14e-02
## age=78, factor(stage_aux)=3      1      0.000  0.7035  7.03e-01  1.11e+00
## age=78, factor(stage_aux)=4      1      0.900  0.1333  4.41e+00  4.78e+00
## age=79, factor(stage_aux)=1 o 2 1      0.000  0.6128  6.13e-01  8.99e-01
## age=79, factor(stage_aux)=3      1      0.911  0.1000  6.58e+00  6.99e+00
## age=81, factor(stage_aux)=1 o 2 2      1.131  0.9261  4.52e-02  6.07e-02
## age=81, factor(stage_aux)=3      1      0.675  0.3625  2.69e-01  3.34e-01
## age=82, factor(stage_aux)=3      1      0.900  0.1333  4.41e+00  4.78e+00
## age=84, factor(stage_aux)=4      1      0.613  0.4005  1.12e-01  1.42e-01
## age=86, factor(stage_aux)=1 o 2 2      1.549  0.4749  2.43e+00  3.15e+00
##
## Chisq= 140  on 62 degrees of freedom, p= 6e-08
```

En ambas pruebas tenemos un p-value de prácticamente cero, por lo que sí hay una diferencia entre las funciones de supervivencia dados estos dos factores. Por lo que, decidimos que en realidad **Sí son significantes**, la razón de que tengamos al cero en el intervalo en 2 de los 3 coeficientes puede deberse a la razón antes mencionada.

Veamos los demás modelos.

Modelo 2: etapas “Avanzada” y “Temprana”

```
confint(modelo_conjunto_2)
```

```
##                                2.5 %      97.5 %
## factor(stage_aux)Temprana(1,2) -1.441913132 -0.31685853
## age                             -0.005684578  0.05141201
```

Ahora ya no tenemos ningún problema con el coeficiente de las etapas dada esta agrupación, es significativo.

Sin embargo, pasó lo mismo en la edad, el coeficiente de la edad no es significativo al 95% de confianza, pero ello se puede deber a la misma razón de no tener muchos datos con fallas y enfrentarnos a muchas censuras, quizás con un nivel de confianza más bajo si sería significativo. Aún así, notemos que está más cargado hacia la derecha que a la izquierda el intervalo, lo cual es bueno.

Como extra: ¿Interactúan?

```
survdif(Surv_general ~ age + factor(stage_aux), rho = 0,
        data = stage_aux_2)
```

```
## Call:
## survdif(formula = Surv_general ~ age + factor(stage_aux), data = stage_aux_2,
##         rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## age=41, factor(stage_aux)=Avanzada(3,4) 1      1  0.1670  4.15e+00  4.26e+00
## age=43, factor(stage_aux)=Temprana(1,2) 1      1  0.4438  6.97e-01  7.17e-01
## age=45, factor(stage_aux)=Temprana(1,2) 1      1  0.3373  1.30e+00  1.33e+00
## age=47, factor(stage_aux)=Temprana(1,2) 2      0  1.2482  1.25e+00  1.31e+00
## age=48, factor(stage_aux)=Avanzada(3,4) 1      0  0.5927  5.93e-01  6.13e-01
## age=48, factor(stage_aux)=Temprana(1,2) 1      0  0.5927  5.93e-01  6.13e-01
## age=49, factor(stage_aux)=Avanzada(3,4) 3      2  0.8456  1.58e+00  1.64e+00
## age=50, factor(stage_aux)=Temprana(1,2) 1      0  1.0725  1.07e+00  1.12e+00
```

```

## age=51, factor(stage_aux)=Avanzada(3,4) 1 0 1.1836 1.18e+00 1.25e+00
## age=51, factor(stage_aux)=Temprana(1,2) 2 0 0.7589 7.59e-01 7.88e-01
## age=52, factor(stage_aux)=Avanzada(3,4) 1 0 0.4830 4.83e-01 4.97e-01
## age=52, factor(stage_aux)=Temprana(1,2) 1 1 0.5695 3.25e-01 3.37e-01
## age=53, factor(stage_aux)=Avanzada(3,4) 1 1 0.2769 1.89e+00 1.93e+00
## age=53, factor(stage_aux)=Temprana(1,2) 2 1 1.2659 5.58e-02 5.85e-02
## age=54, factor(stage_aux)=Avanzada(3,4) 2 1 0.9634 1.39e-03 1.45e-03
## age=56, factor(stage_aux)=Temprana(1,2) 1 0 1.1836 1.18e+00 1.25e+00
## age=57, factor(stage_aux)=Avanzada(3,4) 1 1 0.0801 1.06e+01 1.08e+01
## age=57, factor(stage_aux)=Temprana(1,2) 1 0 0.3373 3.37e-01 3.46e-01
## age=58, factor(stage_aux)=Temprana(1,2) 3 1 2.2098 6.62e-01 7.05e-01
## age=59, factor(stage_aux)=Avanzada(3,4) 1 1 0.6221 2.29e-01 2.37e-01
## age=60, factor(stage_aux)=Avanzada(3,4) 1 1 0.1933 3.37e+00 3.45e+00
## age=60, factor(stage_aux)=Temprana(1,2) 1 1 0.4438 6.97e-01 7.17e-01
## age=61, factor(stage_aux)=Temprana(1,2) 2 0 2.1125 2.11e+00 2.27e+00
## age=62, factor(stage_aux)=Avanzada(3,4) 1 1 0.3217 1.43e+00 1.47e+00
## age=62, factor(stage_aux)=Temprana(1,2) 1 1 0.9956 1.95e-05 2.04e-05
## age=63, factor(stage_aux)=Avanzada(3,4) 1 0 0.5927 5.93e-01 6.13e-01
## age=63, factor(stage_aux)=Temprana(1,2) 3 2 1.2640 4.29e-01 4.51e-01
## age=64, factor(stage_aux)=Avanzada(3,4) 1 1 0.2206 2.75e+00 2.82e+00
## age=64, factor(stage_aux)=Temprana(1,2) 3 1 1.5767 2.11e-01 2.22e-01
## age=65, factor(stage_aux)=Avanzada(3,4) 4 3 1.9577 5.55e-01 5.93e-01
## age=66, factor(stage_aux)=Avanzada(3,4) 1 0 0.5927 5.93e-01 6.13e-01
## age=66, factor(stage_aux)=Temprana(1,2) 2 0 1.6177 1.62e+00 1.71e+00
## age=67, factor(stage_aux)=Temprana(1,2) 2 1 1.2663 5.60e-02 5.85e-02
## age=68, factor(stage_aux)=Avanzada(3,4) 2 2 1.3905 2.67e-01 2.82e-01
## age=68, factor(stage_aux)=Temprana(1,2) 3 1 2.8489 1.20e+00 1.31e+00
## age=69, factor(stage_aux)=Avanzada(3,4) 3 1 2.1121 5.86e-01 6.22e-01
## age=70, factor(stage_aux)=Avanzada(3,4) 1 1 0.7811 6.14e-02 6.35e-02
## age=70, factor(stage_aux)=Temprana(1,2) 2 1 1.1384 1.68e-02 1.76e-02
## age=71, factor(stage_aux)=Avanzada(3,4) 3 3 0.5959 9.70e+00 1.00e+01
## age=71, factor(stage_aux)=Temprana(1,2) 1 0 0.3063 3.06e-01 3.14e-01
## age=72, factor(stage_aux)=Avanzada(3,4) 1 1 0.2769 1.89e+00 1.93e+00
## age=72, factor(stage_aux)=Temprana(1,2) 1 0 0.4830 4.83e-01 4.97e-01
## age=73, factor(stage_aux)=Temprana(1,2) 2 0 2.2561 2.26e+00 2.45e+00
## age=74, factor(stage_aux)=Avanzada(3,4) 2 1 0.5857 2.93e-01 3.03e-01
## age=74, factor(stage_aux)=Temprana(1,2) 1 1 0.7356 9.50e-02 9.83e-02
## age=75, factor(stage_aux)=Temprana(1,2) 1 1 0.6939 1.35e-01 1.40e-01
## age=76, factor(stage_aux)=Avanzada(3,4) 1 1 0.0682 1.27e+01 1.30e+01
## age=76, factor(stage_aux)=Temprana(1,2) 1 1 0.3882 9.64e-01 9.88e-01
## age=77, factor(stage_aux)=Temprana(1,2) 3 2 1.6624 6.86e-02 7.25e-02
## age=78, factor(stage_aux)=Avanzada(3,4) 2 1 1.3250 7.97e-02 8.40e-02
## age=79, factor(stage_aux)=Avanzada(3,4) 1 1 0.1044 7.69e+00 7.84e+00
## age=79, factor(stage_aux)=Temprana(1,2) 1 0 0.9289 9.29e-01 9.69e-01
## age=81, factor(stage_aux)=Avanzada(3,4) 1 1 0.4438 6.97e-01 7.17e-01
## age=81, factor(stage_aux)=Temprana(1,2) 2 2 1.2249 4.90e-01 5.15e-01
## age=82, factor(stage_aux)=Avanzada(3,4) 1 1 0.1414 5.21e+00 5.34e+00
## age=84, factor(stage_aux)=Avanzada(3,4) 1 1 0.5043 4.87e-01 5.02e-01
## age=86, factor(stage_aux)=Temprana(1,2) 2 2 0.6151 3.12e+00 3.22e+00
##
## Chisq= 104 on 56 degrees of freedom, p= 1e-04
survdifff(Surv_general ~ age + factor(stage_aux), rho = 1,
          data = stage_aux_2)

```

```
## Call:
## survdiff(formula = Surv_general ~ age + factor(stage_aux), data = stage_aux_2,
##          rho = 1)
##
##
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
## age=41, factor(stage_aux)=Avanzada(3,4)	1	0.867	0.1556	3.25e+00	3.57e+00
## age=43, factor(stage_aux)=Temprana(1,2)	1	0.675	0.3625	2.69e-01	3.34e-01
## age=45, factor(stage_aux)=Temprana(1,2)	1	0.722	0.2892	6.48e-01	7.66e-01
## age=47, factor(stage_aux)=Temprana(1,2)	2	0.000	0.9391	9.39e-01	1.27e+00
## age=48, factor(stage_aux)=Avanzada(3,4)	1	0.000	0.4526	4.53e-01	5.94e-01
## age=48, factor(stage_aux)=Temprana(1,2)	1	0.000	0.4526	4.53e-01	5.94e-01
## age=49, factor(stage_aux)=Avanzada(3,4)	3	1.844	0.6799	1.99e+00	2.50e+00
## age=50, factor(stage_aux)=Temprana(1,2)	1	0.000	0.6664	6.66e-01	1.02e+00
## age=51, factor(stage_aux)=Avanzada(3,4)	1	0.000	0.7035	7.03e-01	1.11e+00
## age=51, factor(stage_aux)=Temprana(1,2)	2	0.000	0.6379	6.38e-01	7.79e-01
## age=52, factor(stage_aux)=Avanzada(3,4)	1	0.000	0.3875	3.87e-01	4.87e-01
## age=52, factor(stage_aux)=Temprana(1,2)	1	0.599	0.4396	5.81e-02	7.57e-02
## age=53, factor(stage_aux)=Avanzada(3,4)	1	0.778	0.2444	1.16e+00	1.34e+00
## age=53, factor(stage_aux)=Temprana(1,2)	2	0.844	0.8442	6.84e-08	9.75e-08
## age=54, factor(stage_aux)=Avanzada(3,4)	2	0.711	0.7656	3.92e-03	5.01e-03
## age=56, factor(stage_aux)=Temprana(1,2)	1	0.000	0.7035	7.03e-01	1.11e+00
## age=57, factor(stage_aux)=Avanzada(3,4)	1	0.933	0.0778	9.41e+00	9.90e+00
## age=57, factor(stage_aux)=Temprana(1,2)	1	0.000	0.2892	2.89e-01	3.42e-01
## age=58, factor(stage_aux)=Temprana(1,2)	3	0.711	1.5029	4.17e-01	6.00e-01
## age=59, factor(stage_aux)=Avanzada(3,4)	1	0.547	0.4687	1.32e-02	1.75e-02
## age=60, factor(stage_aux)=Avanzada(3,4)	1	0.844	0.1778	2.50e+00	2.78e+00
## age=60, factor(stage_aux)=Temprana(1,2)	1	0.675	0.3625	2.69e-01	3.34e-01
## age=61, factor(stage_aux)=Temprana(1,2)	2	0.000	1.3163	1.32e+00	2.04e+00
## age=62, factor(stage_aux)=Avanzada(3,4)	1	0.733	0.2779	7.46e-01	8.77e-01
## age=62, factor(stage_aux)=Temprana(1,2)	1	0.387	0.6386	9.90e-02	1.48e-01
## age=63, factor(stage_aux)=Avanzada(3,4)	1	0.000	0.4526	4.53e-01	5.94e-01
## age=63, factor(stage_aux)=Temprana(1,2)	3	1.355	1.0313	1.02e-01	1.28e-01
## age=64, factor(stage_aux)=Avanzada(3,4)	1	0.811	0.2000	1.87e+00	2.10e+00
## age=64, factor(stage_aux)=Temprana(1,2)	3	0.800	1.2016	1.34e-01	1.79e-01
## age=65, factor(stage_aux)=Avanzada(3,4)	4	2.352	1.3486	7.46e-01	1.07e+00
## age=66, factor(stage_aux)=Avanzada(3,4)	1	0.000	0.4526	4.53e-01	5.94e-01
## age=66, factor(stage_aux)=Temprana(1,2)	2	0.000	1.1074	1.11e+00	1.59e+00
## age=67, factor(stage_aux)=Temprana(1,2)	2	0.409	0.9020	2.70e-01	3.71e-01
## age=68, factor(stage_aux)=Avanzada(3,4)	2	1.156	0.8924	7.78e-02	1.13e-01
## age=68, factor(stage_aux)=Temprana(1,2)	3	0.361	1.8226	1.17e+00	1.80e+00
## age=69, factor(stage_aux)=Avanzada(3,4)	3	0.756	1.4389	3.25e-01	4.64e-01
## age=70, factor(stage_aux)=Avanzada(3,4)	1	0.473	0.5483	1.03e-02	1.44e-02
## age=70, factor(stage_aux)=Temprana(1,2)	2	0.638	0.8739	6.40e-02	8.44e-02
## age=71, factor(stage_aux)=Avanzada(3,4)	3	2.593	0.4986	8.80e+00	1.07e+01
## age=71, factor(stage_aux)=Temprana(1,2)	1	0.000	0.2667	2.67e-01	3.12e-01
## age=72, factor(stage_aux)=Avanzada(3,4)	1	0.778	0.2444	1.16e+00	1.34e+00
## age=72, factor(stage_aux)=Temprana(1,2)	1	0.000	0.3875	3.87e-01	4.87e-01
## age=73, factor(stage_aux)=Temprana(1,2)	2	0.000	1.3699	1.37e+00	2.17e+00
## age=74, factor(stage_aux)=Avanzada(3,4)	2	0.800	0.5115	1.63e-01	1.92e-01
## age=74, factor(stage_aux)=Temprana(1,2)	1	0.494	0.5268	2.06e-03	2.85e-03
## age=75, factor(stage_aux)=Temprana(1,2)	1	0.514	0.5062	1.07e-04	1.45e-04
## age=76, factor(stage_aux)=Avanzada(3,4)	1	0.944	0.0667	1.16e+01	1.21e+01
## age=76, factor(stage_aux)=Temprana(1,2)	1	0.687	0.3250	4.03e-01	4.88e-01
## age=77, factor(stage_aux)=Temprana(1,2)	3	1.374	1.1864	2.96e-02	4.14e-02

```
## age=78, factor(stage_aux)=Avanzada(3,4) 2    0.900    0.8368    4.77e-03    7.04e-03
## age=79, factor(stage_aux)=Avanzada(3,4) 1    0.911    0.1000    6.58e+00    6.99e+00
## age=79, factor(stage_aux)=Temprana(1,2) 1    0.000    0.6128    6.13e-01    8.99e-01
## age=81, factor(stage_aux)=Avanzada(3,4) 1    0.675    0.3625    2.69e-01    3.34e-01
## age=81, factor(stage_aux)=Temprana(1,2) 2    1.131    0.9261    4.52e-02    6.07e-02
## age=82, factor(stage_aux)=Avanzada(3,4) 1    0.900    0.1333    4.41e+00    4.78e+00
## age=84, factor(stage_aux)=Avanzada(3,4) 1    0.613    0.4005    1.12e-01    1.42e-01
## age=86, factor(stage_aux)=Temprana(1,2) 2    1.549    0.4749    2.43e+00    3.15e+00
##
## Chisq= 89.5 on 56 degrees of freedom, p= 0.003
```

En ambos tenemos un p-value muy grande, por lo que podríamos decir que sí hay una diferencia en la supervivencia del cáncer bajo estos factores, por lo que los dejaremos; es decir, consideramos que **También son significativos bajo esta clasificación** ### Modelo 3: Con grupo de edad de 40 a 70 y de 71 en adelante; etapas 1 y 2 agrupadas

```
confint(modelo_conjunto_3)
```

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

Agrupando las etapas 1 y 2, así como la edad, tenemos ambos parámetros como significativos, y ello hace mucho sentido intuitivamente también, ya que las etapas 1 y 2 tienen, naturalmente, mucho menos riesgo que la 3 y 4 que son las más avanzadas, y en el grupo de edad evidentemente no tienen las mismas defensas los adultos que las personas de la tercera edad.

¿Interactúan?

```
survdifff(Surv_general ~ Grupo+ factor(stage_aux), rho = 0,
          data = stage_aux_2)
```

```
## Call:
## survdifff(formula = Surv_general ~ Grupo + factor(stage_aux),
##           data = stage_aux_2, rho = 0)
##
##                                N Observed Expected (O-E)^2/E
## Grupo=(40,70], factor(stage_aux)=Avanzada(3,4) 27      17     13.38    0.98132
## Grupo=(40,70], factor(stage_aux)=Temprana(1,2) 35      13     23.28    4.54169
## Grupo=(70,Inf], factor(stage_aux)=Avanzada(3,4) 13      11      4.05   11.95601
## Grupo=(70,Inf], factor(stage_aux)=Temprana(1,2) 15       9      9.29    0.00933
##
##                                (O-E)^2/V
## Grupo=(40,70], factor(stage_aux)=Avanzada(3,4)    1.3644
## Grupo=(40,70], factor(stage_aux)=Temprana(1,2)    8.6837
## Grupo=(70,Inf], factor(stage_aux)=Avanzada(3,4)   13.4061
## Grupo=(70,Inf], factor(stage_aux)=Temprana(1,2)    0.0117
##
## Chisq= 18.1 on 3 degrees of freedom, p= 4e-04
```

```
survdifff(Surv_general ~ Grupo+ factor(stage_aux), rho = 1,
          data = stage_aux_2)
```

```
## Call:
## survdifff(formula = Surv_general ~ Grupo + factor(stage_aux),
##           data = stage_aux_2, rho = 1)
##
##                                N Observed Expected (O-E)^2/E
## Grupo=(40,70], factor(stage_aux)=Avanzada(3,4) 27    12.81     9.72    0.976
```

```
## Grupo=(40,70], factor(stage_aux)=Temprana(1,2) 35      8.18    16.38    4.111
## Grupo=(70,Inf], factor(stage_aux)=Avanzada(3,4) 13      9.11     3.15    11.260
## Grupo=(70,Inf], factor(stage_aux)=Temprana(1,2) 15      5.75     6.58     0.106
##
##                                     (O-E)^2/V
## Grupo=(40,70], factor(stage_aux)=Avanzada(3,4)    1.768
## Grupo=(40,70], factor(stage_aux)=Temprana(1,2)    10.133
## Grupo=(70,Inf], factor(stage_aux)=Avanzada(3,4)    15.529
## Grupo=(70,Inf], factor(stage_aux)=Temprana(1,2)     0.175
##
##  Chisq= 21.2  on 3 degrees of freedom, p= 1e-04
```

En ambos casos tenemos un p-balue pequeñísimo, por lo que estos factores sí tienen un efecto en la supervivencia al cáncer.

Consideramos que **También son significativos bajo esta clasificación** # 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:etapa 1 y 2 agrupadas

Utilizamos la funcion `cox.zph` para hacer el test al modelo, donde:

Ho: los riesgos son proporcionales vs H1: los riesgos no son proporcionales .

```
sup_1 = cox.zph(modelo_conjunto_1)
sup_1
```

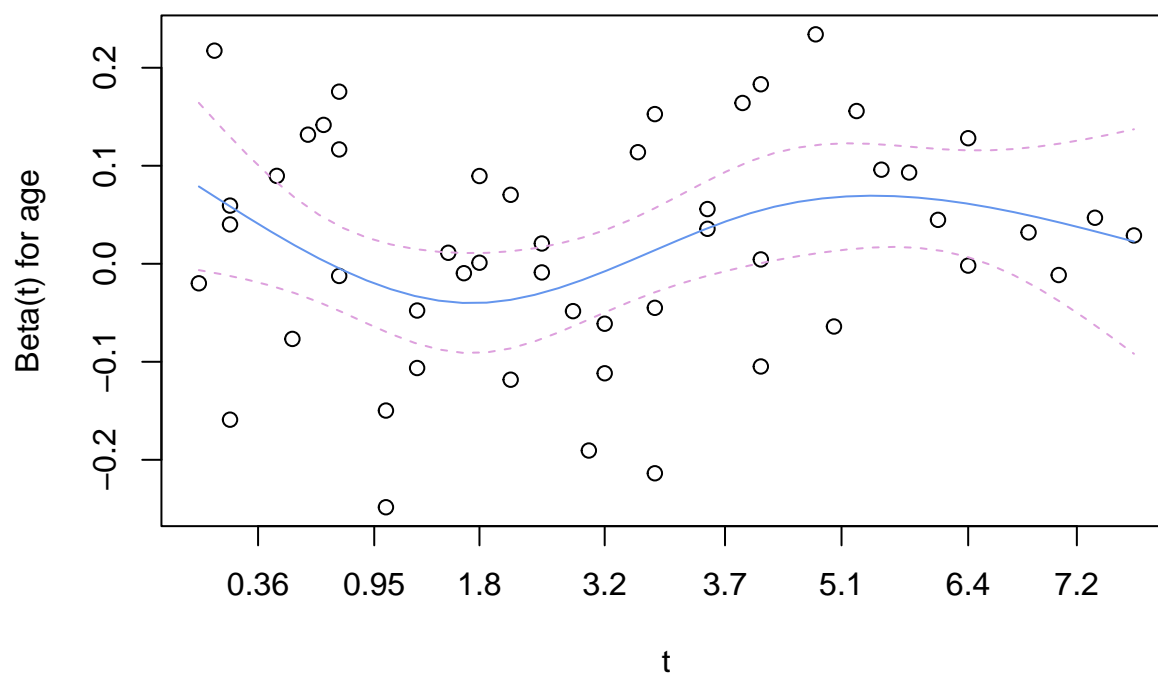
```
##               chisq df    p
## factor(stage_aux) 3.60  2 0.17
## age               1.03  1 0.31
## GLOBAL            4.87  3 0.18
```

Ya que el p-value > alpha para cada una de las covariables se concluye que no existen pruebas suficientes para rechazar Ho. Es decir el cociente de riesgos es independiente del tiempo.

Verifiquemos gráficamente los residuos Cox-Snell:

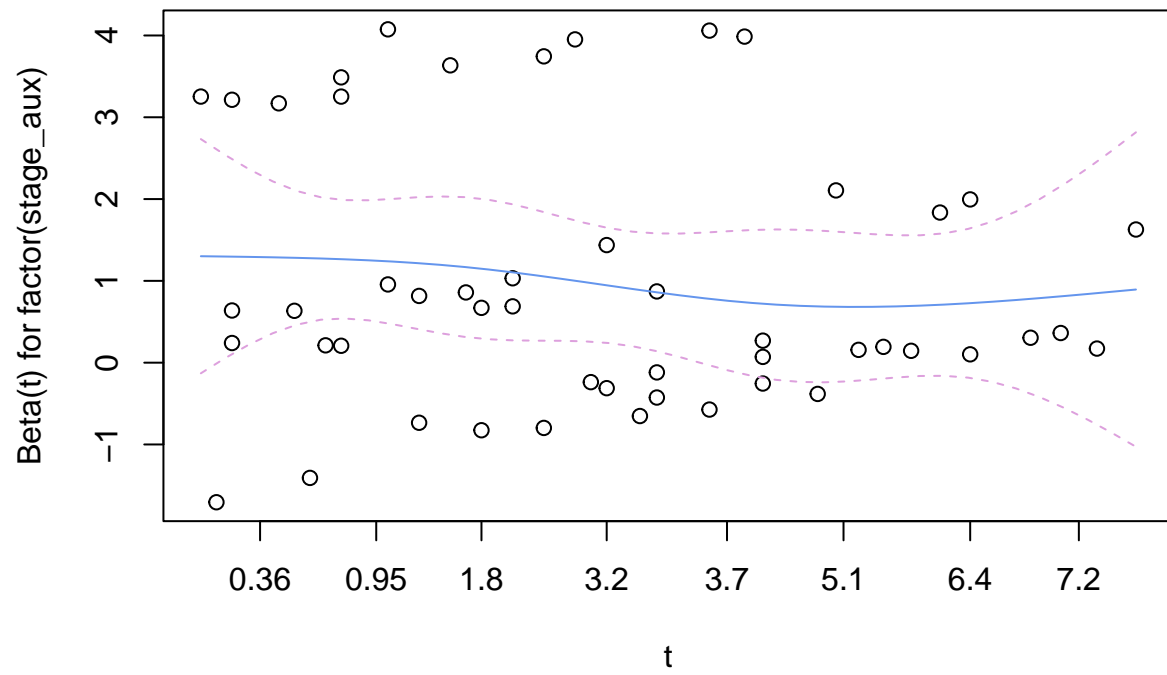
```
plot(sup_1, var = "age", xlab = "t", main = "Variable 'edad' en el tiempo",
     col = c("cornflowerblue", "plum"))
```

Variable 'edad' en el tiempo

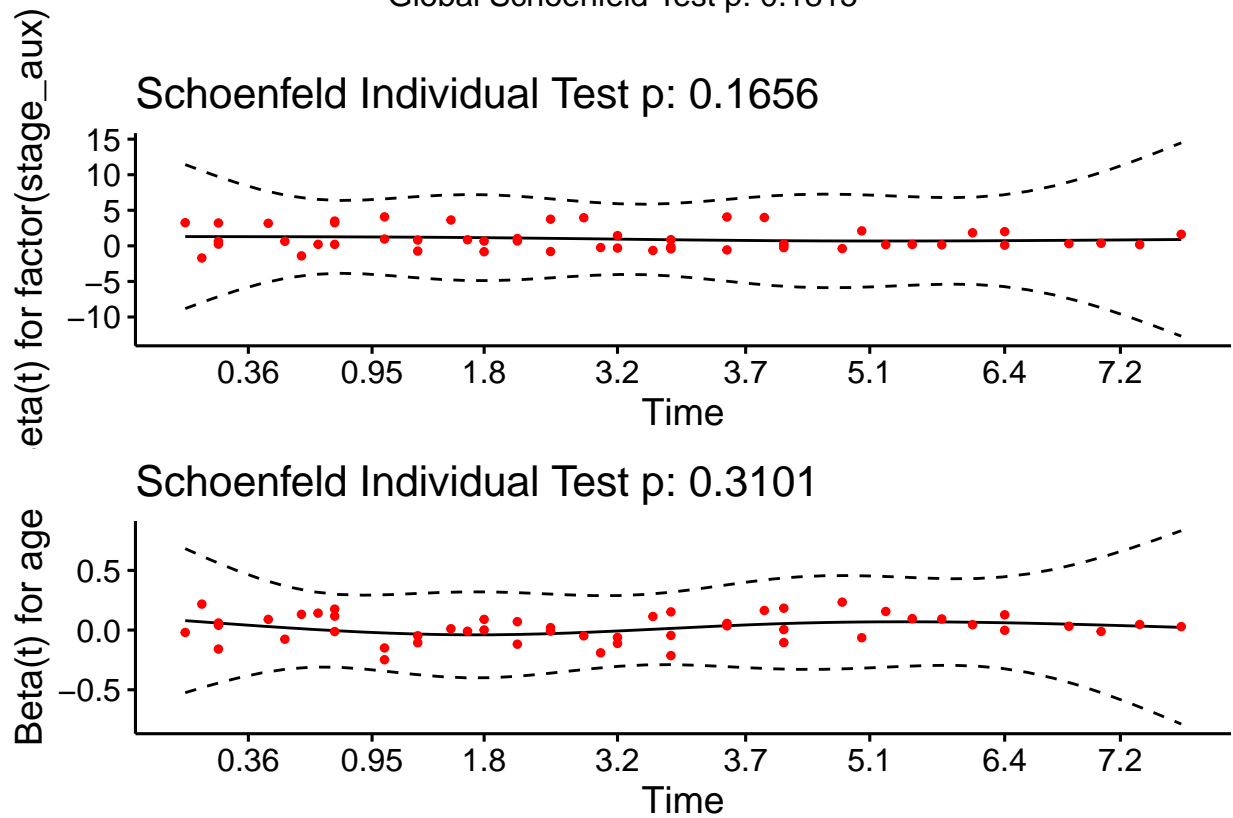


```
# Y, finalmente,  
plot(sup_1, var = "factor(stage_aux)", xlab = "t",  
      main = "Etapa", col = c("cornflowerblue", "plum"))  
ggcoxzph(sup_1) #verificacion grafica
```


Etapa



Global Schoenfeld Test p: 0.1815



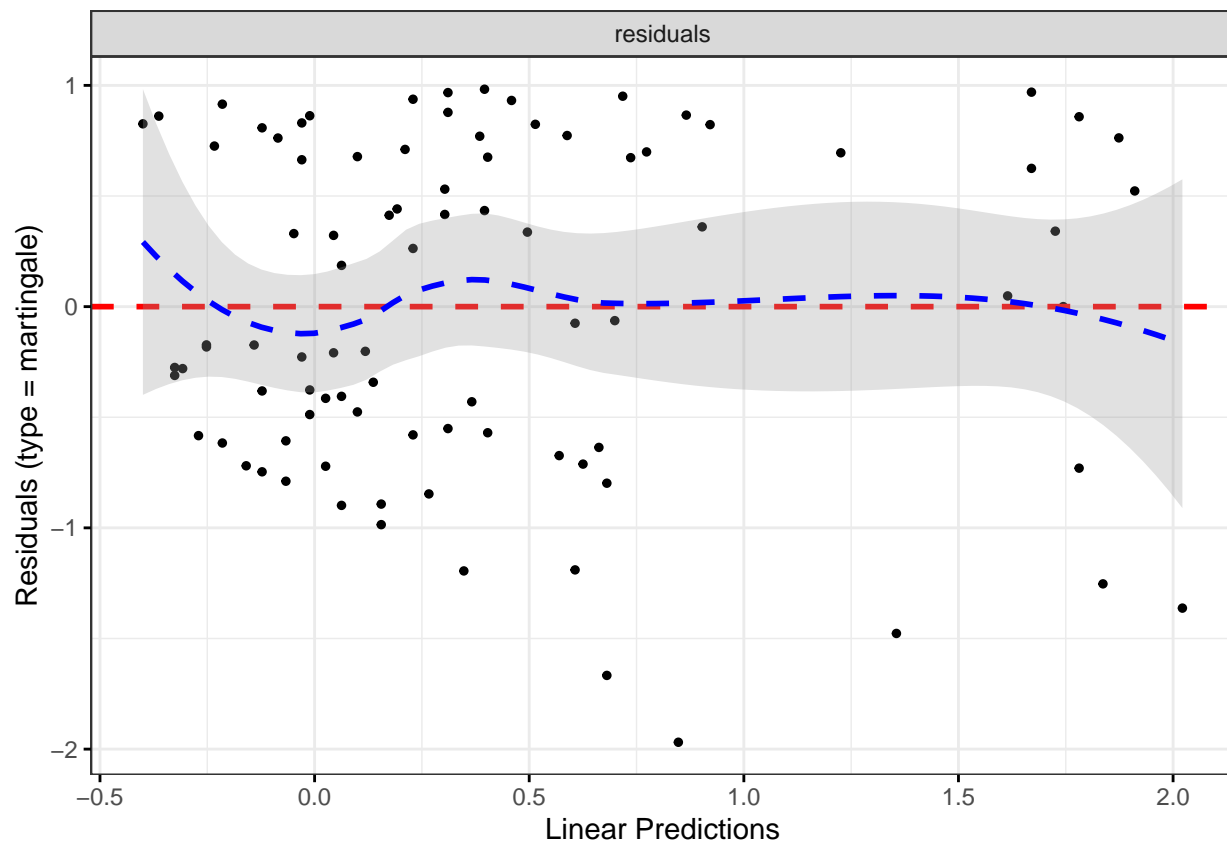
Vemos que los residuos se agrupan de manera aleatoria a ambos lados del valor 0 del eje y, por lo que se cumple el supuesto, inclusive la linea azul se pega bastante al cero. Además de que tenemos el comportamiento deseado, y de hecho pasamos la prueba de Schoenfeld con un p-value mayor a 0.05 en los 3 casos.

Todo bien hasta aquí

Veamos los martingala:

```
ggcoxdiagnostics(modelo_conjunto_1, type = "martingale")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

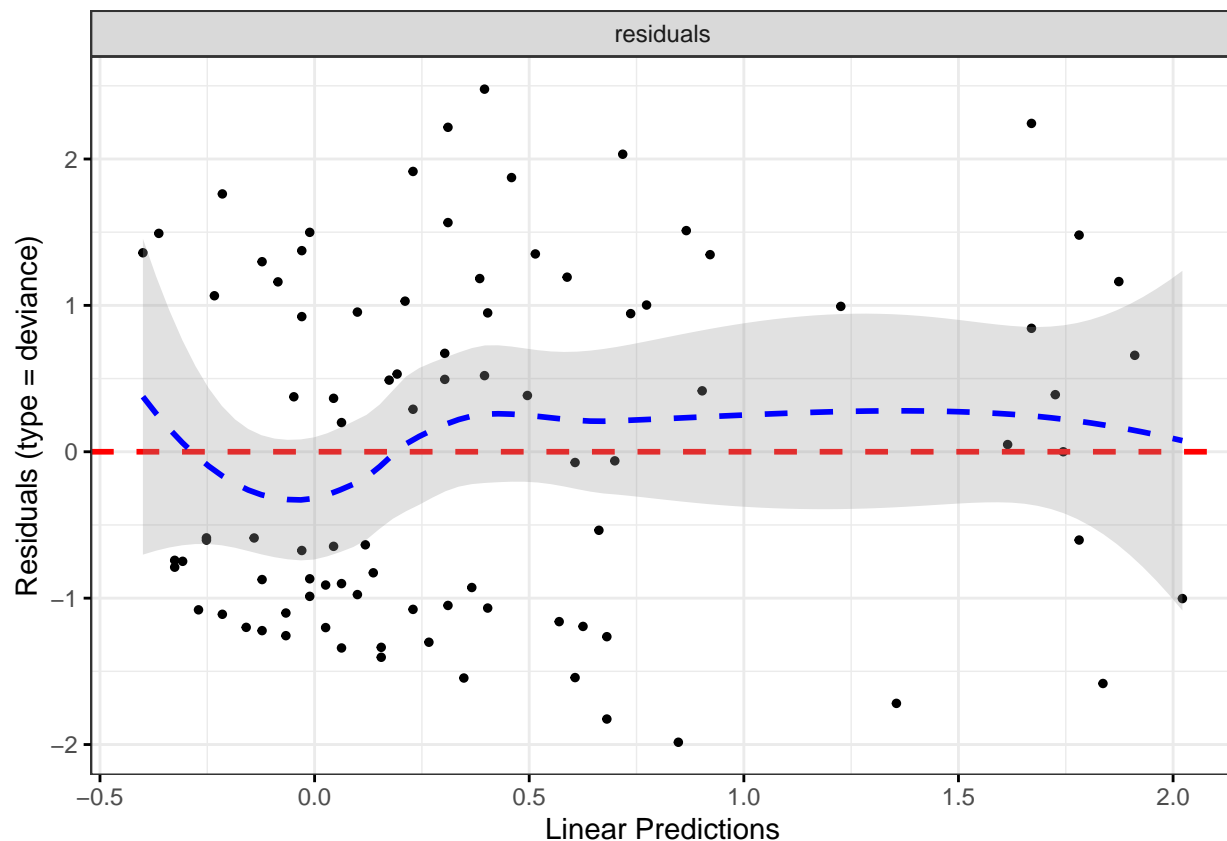


Vemos que, de igual manera hay algunos puntos (varios) que salen de los intervalos de confianza, pero la línea azul se pega bastante al cero, por lo que podemos considerar que, salvo el inicio y final, se ve bien.

Veamos los de devianza:

```
ggcoxdiagnostics(modelo_conjunto_1, type = "deviance")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



De igual manera la línea se pega bastante al cero, y parece tener un comportamiento aleatorio donde no tenemos algún comportamiento en particular

```
dev_1 = residuals(modelo_conjunto_1, type = "deviance")
min(dev_1)
```

```
## [1] -1.984415
```

```
max(dev_1)
```

```
## [1] 2.477164
```

```
summary(dev_1)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -1.98441 -1.03815 -0.03101  0.02120  1.02222  2.47716
```

```
# Casos extremos:
```

```
dev_1[abs(dev_1) >= 1] # Posicion
```

```
##      1      2      3      5      7      9     10     11
## 1.914951 1.761420 1.491927 1.298335 1.028863 1.359340 1.160620 1.065755
##      21     25     26     27     29     30     31     32
## -1.076695 -1.301158 -1.101517 -1.201166 -1.336053 -1.199524 -1.403877 -1.221931
##      33     34     35     36     48     49     50     51
## -1.340402 2.477164 1.498997 1.373786 -1.079918 -1.110317 -1.256328 2.216586
##      52     53     54     55     56     57     58     59
## 2.032652 1.873107 1.510816 1.346538 1.565709 1.351468 1.192725 1.002299
##      61     65     66     67     69     70     73     75
## 1.183295 -1.193130 -1.067731 -1.160463 -1.049909 -1.263364 -1.542679 -1.984415
```

```
##          76          77          78          79          80          87          89          90
## -1.825841 -1.545695  2.243209  1.480122  1.162577 -1.583070 -1.002884 -1.718830
```

```
# Casos muy extremos:
dev_1[abs(dev_1) >= 2] # Posicion
```

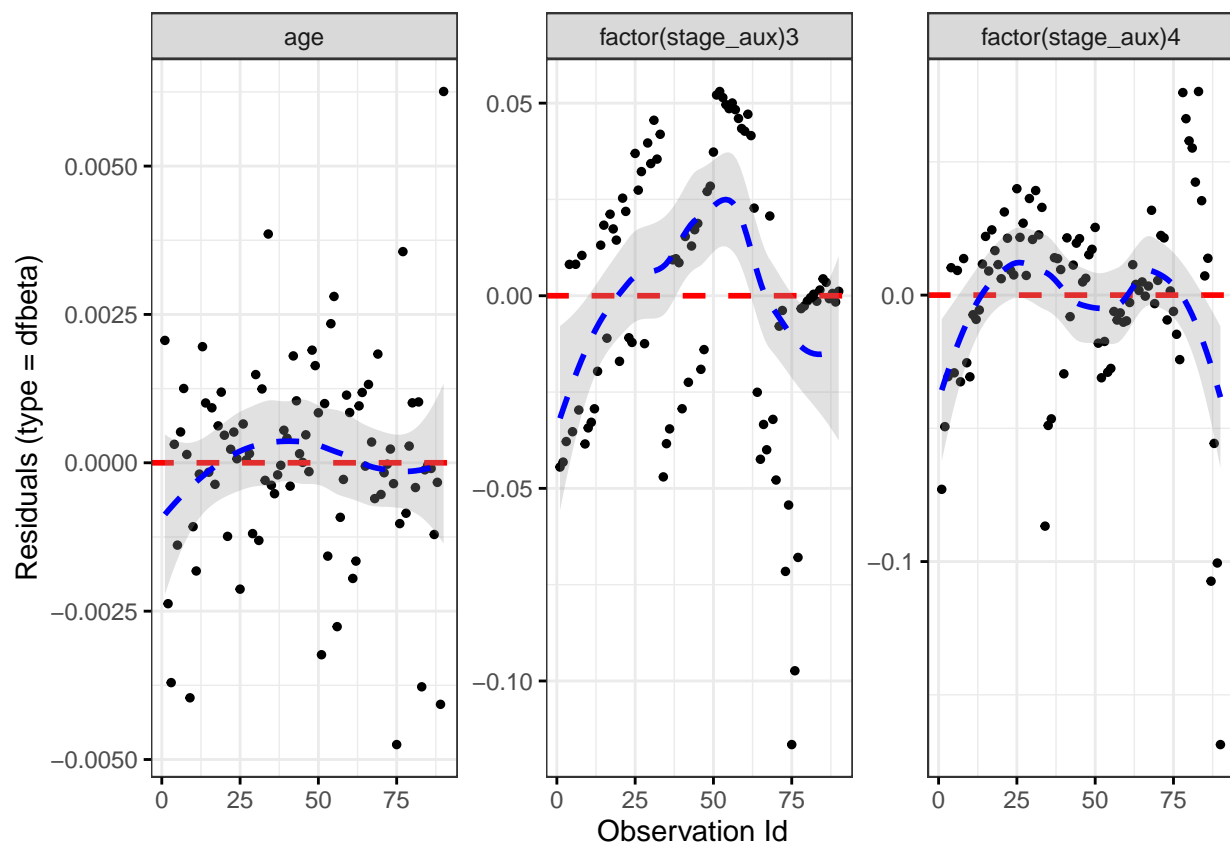
```
##          34          51          52          78
## 2.477164 2.216586 2.032652 2.243209
```

La media de los residuos es bastante cercana al cero, además de que si bien tenemos varios datos extremos, solo 4 sobrepasan las 2 unidades, pero ninguna supera los 2.5; estos pueden ser considerados datos atípicos

Veamos los DfBeta:

```
ggcoxdiagnostics(modelo_conjunto_1, type = "dfbeta")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Aquí la línea azul no se pega tanto al cero, en el caso de la etaoa 3; en la edad parece ser que sí, pero en la etapa 4 tenemos problemas al inicio y al final

Parece ser que tenemos algunos datos atípicos, tal como nos lo sugiere el análisis de devianza

Modelo 2: etapas “Avanzada” y “Temprana”

Utilizamos la funcion `cox.zph` para hacer el test al modelo, donde:

Ho: los riesgos son proporcionales vs H1: los riesgos no son proporcionales .

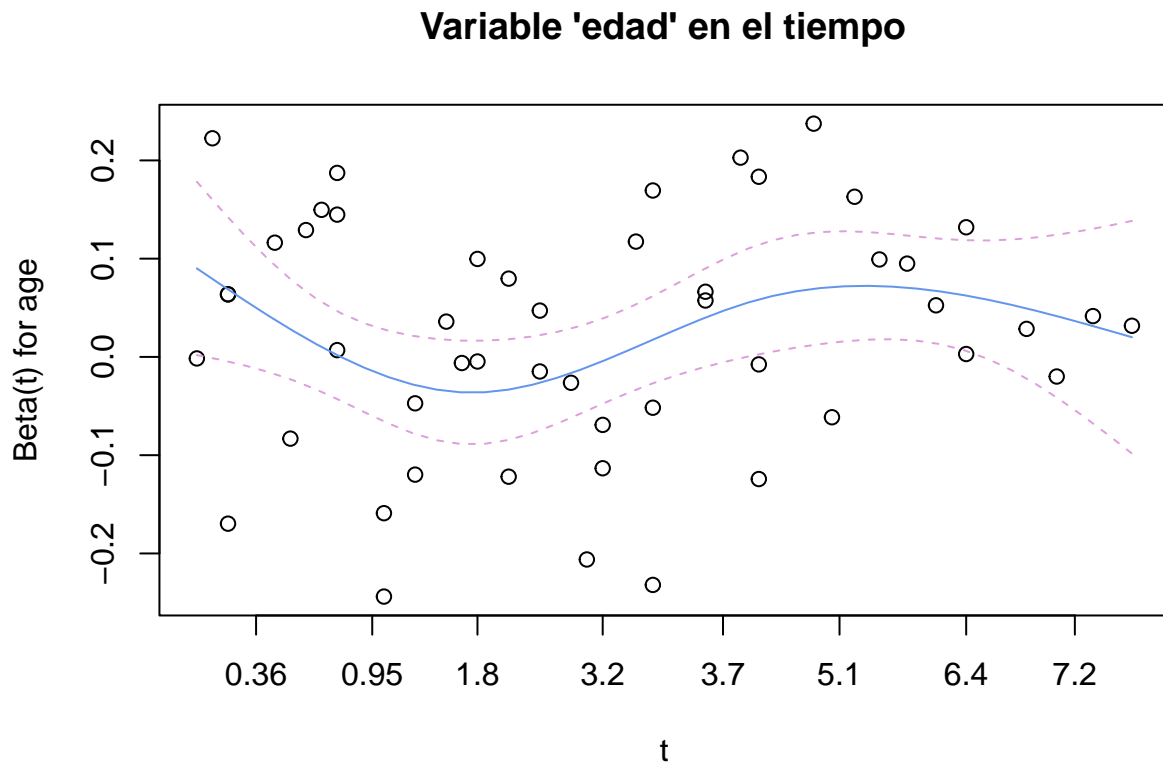
```
sup_2 = cox.zph(modelo_conjunto_2)
sup_2
```

```
##               chisq df      p
## factor(stage_aux) 5.181  1 0.023
## age               0.705  1 0.401
## GLOBAL            6.227  2 0.044
```

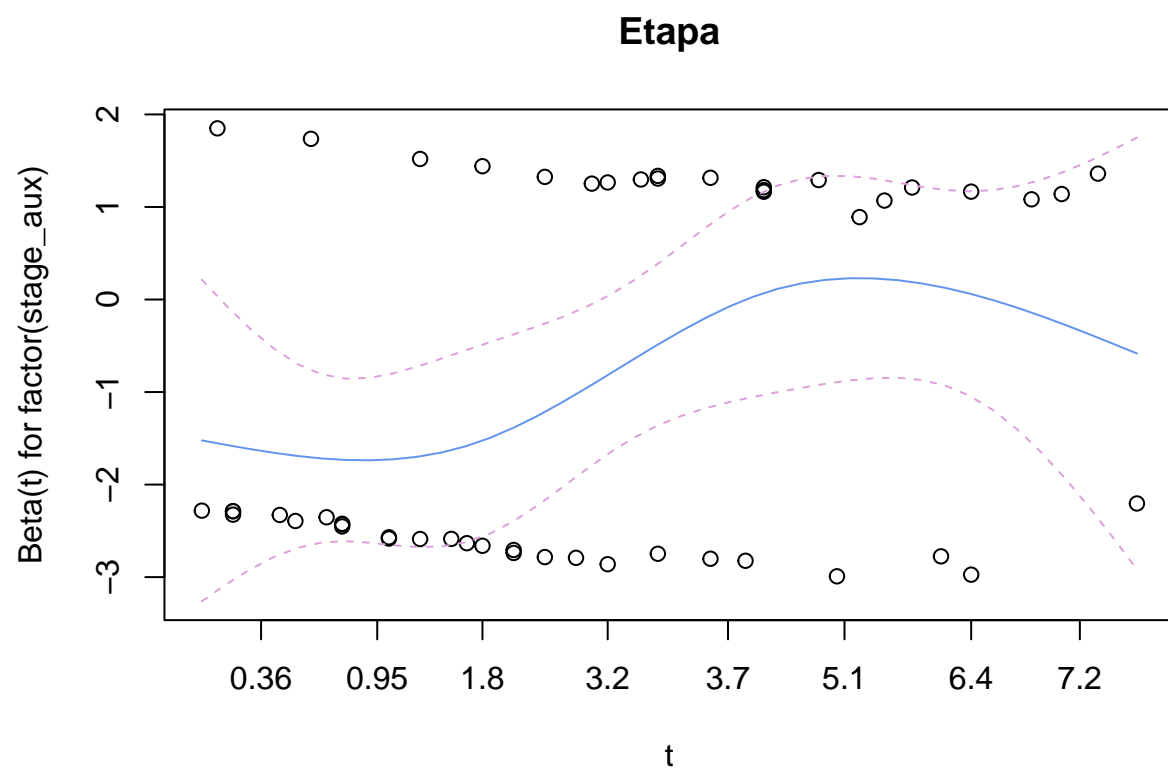
Lamentablemente, para las etapas no se pasa la prueba, tampoco de manera global, por lo que hay evidencia significativa estadísticamente como para decir que los riesgos no son proporcionales. **No es válido el supuesto en este modelo**

Verifiquemos gráficamente los residuos de Cox - Snell:

```
plot(sup_2, var = "age", xlab = "t", main = "Variable 'edad' en el tiempo",
      col = c("cornflowerblue", "plum"))
```

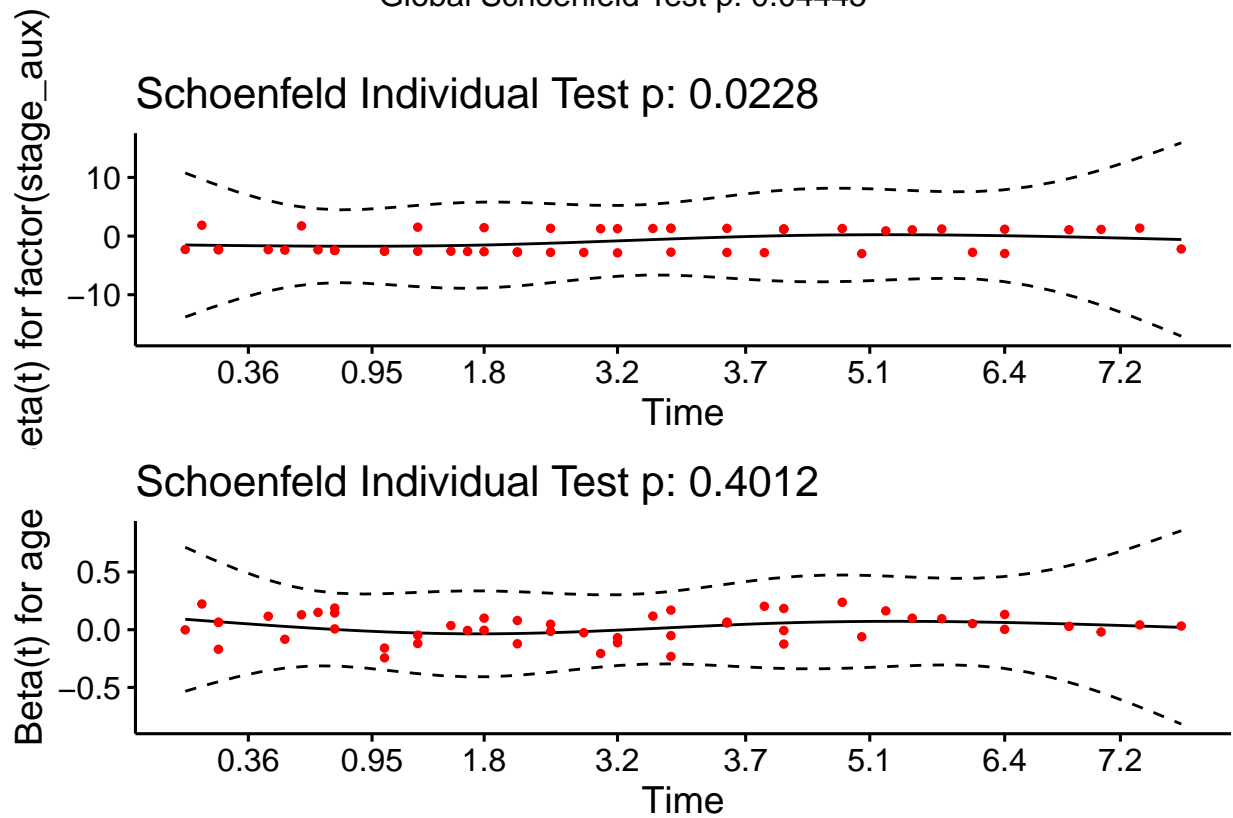


```
# Y, finalmente,
plot(sup_2, var = "factor(stage_aux)", xlab = "t",
      main = "Etapa", col = c("cornflowerblue", "plum"))
```



```
ggcoxzph(sup_2)#verificacion grafica
```

Global Schoenfeld Test p: 0.04445

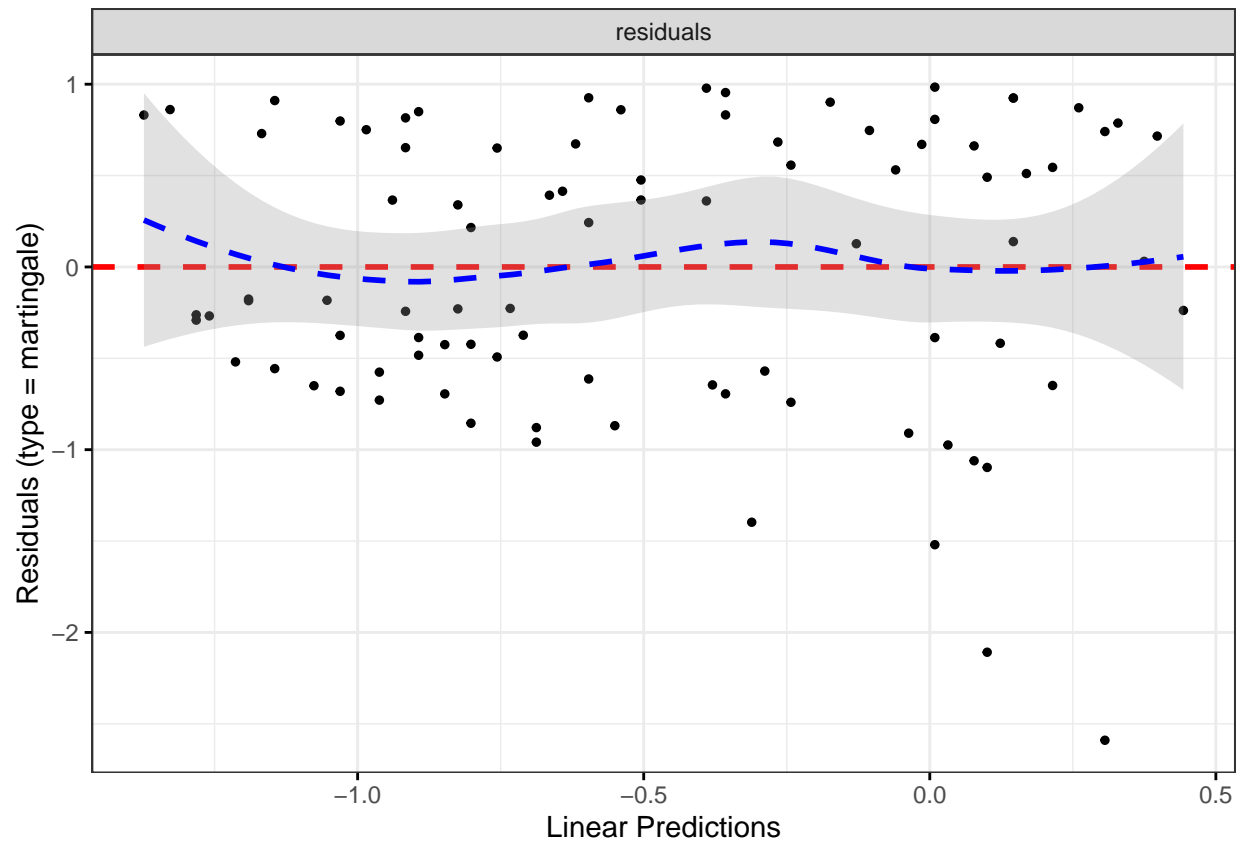


Vemos que parece seguir un tipo de patrón para la etapa, no cumple con la característica de tener un comportamiento aleatorio alrededor del eje y; inclusive no se pasa el test de manera conjunta. Además, en el factor la línea azul no se pega mucho al cero.

Veamos los martingala:

```
ggcoxdiagnostics(modelo_conjunto_2, type = "martingale")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

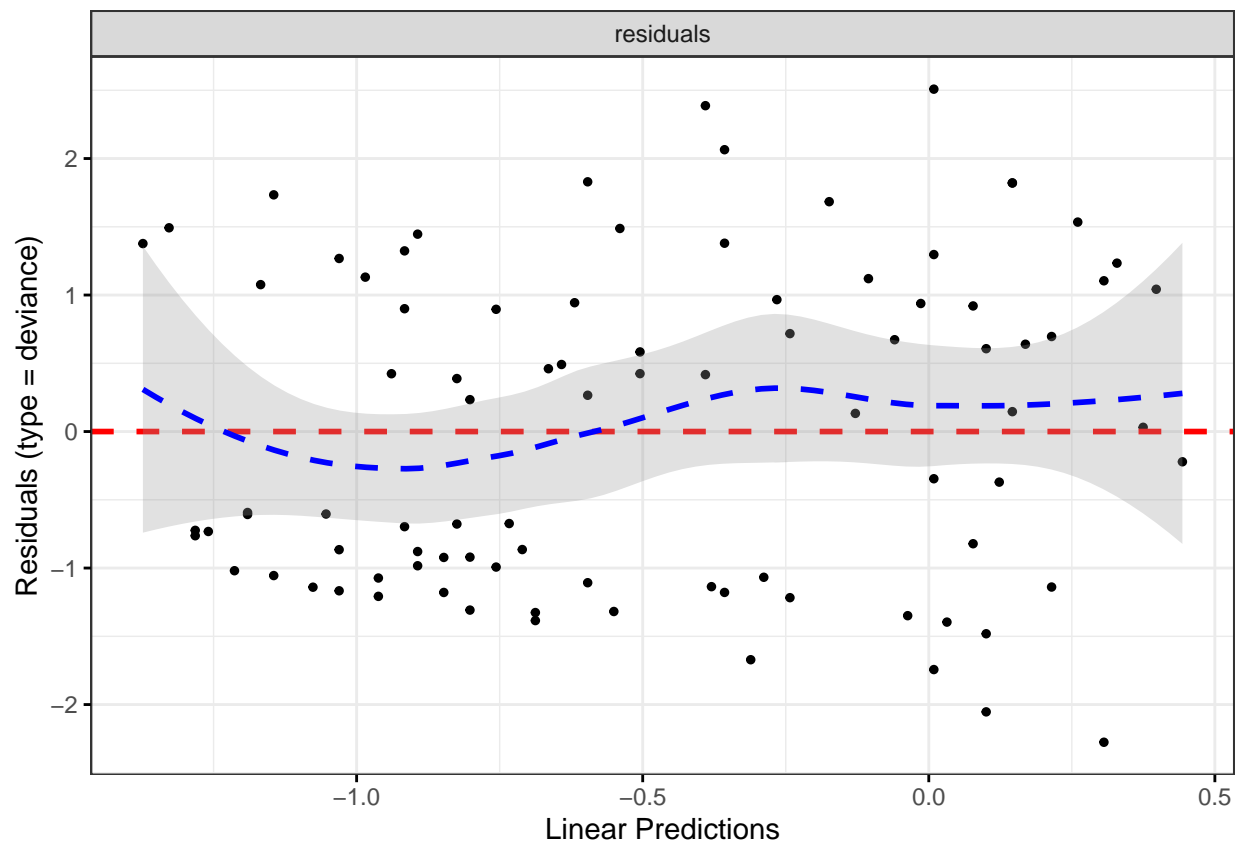



Aunque vario datos se salen de los intervalos de confianza, observamos que la línea azul se pega bastante al cero, por lo que vamos bien aquí.

Veamos los de devianza:

```
ggcoxdiagnostics(modelo_conjunto_2, type = "deviance")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
dev_2 = residuals(modelo_conjunto_2, type = "deviance")
min(dev_2)
```

```
## [1] -2.276017
```

```
max(dev_2)
```

```
## [1] 2.508233
```

```
summary(dev_2)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -2.27602 -1.04611  0.08199  0.02005  1.02331  2.50823
```

```
# Casos extremos:
```

```
dev_2[abs(dev_2) >= 1] # Posicion
```

```
##      1      2      3      5      9     10     11     21
## 1.829460 1.733993 1.492622 1.267918 1.377067 1.131043 1.076162 -1.107347
##      25     26     27     29     30     31     32     33
## -1.318227 -1.073060 -1.178757 -1.326020 -1.140233 -1.384826 -1.166604 -1.307890
##      34     35     36     48     49     50     51     52
## 2.387152 1.446057 1.323522 -1.019435 -1.055004 -1.207307 2.064627 1.821106
##      53     54     55     56     57     64     65     66
## 1.683851 1.233604 1.042351 1.379580 1.120216 -1.067571 -1.395934 -1.216991
##      67     69     70     73     75     76     77     78
## -1.348871 -1.178547 -1.481300 -1.743570 -2.276017 -2.053488 -1.671576 2.508233
##      79     80     81     82     83     87     90
## 1.821106 1.534414 1.296556 1.104444 1.487576 -1.139028 -1.136315
```

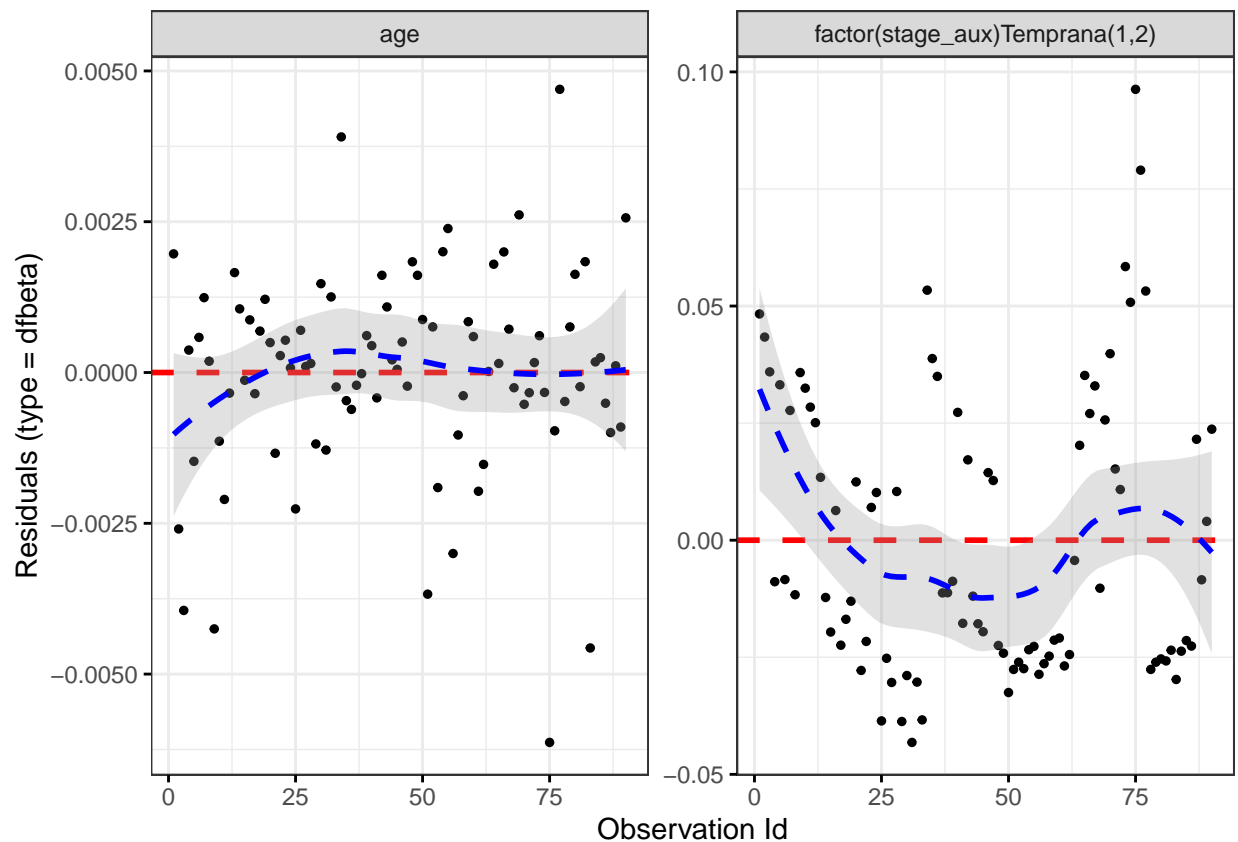
```
dev_2[abs(dev_2) >= 2] # Posicion
```

```
##          34          51          75          76          78
## 2.387152 2.064627 -2.276017 -2.053488 2.508233
```

La misma dinámica: Parece tener un comportamiento algo aleatorio, además de que tenemos a la línea azul muy pegada al cero. Si bien tenemos algunos datos que se salen de los intervalos, podemos ver que solo 5 datos son mayores a 2, y solo uno de ellos es mayor a 5, siendo este un dato que puede estar afectando demasiado nuestro modelo al ser tan atípico

```
ggcoxdiagnostics(modelo_conjunto_2, type = "dfbeta")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



La línea azul de la edad sí se pega bastante al cero, sin embargo no es el caso para las etapas, aunque el problema está más en el inicio y final.

Para la etapa tenemos datos bastante salidos de los intervalos de confianza, que pueden ser los que nos está indicando el análisis de devianza

Modelo 3: etapas 1 y 2 agrupadas

Utilizamos la función `cox.zph` para hacer el test al modelo, donde:

H_0 : los riesgos son proporcionales vs H_1 : los riesgos no son proporcionales .

```
sup_3 = cox.zph(modelo_conjunto_3)
sup_3
```

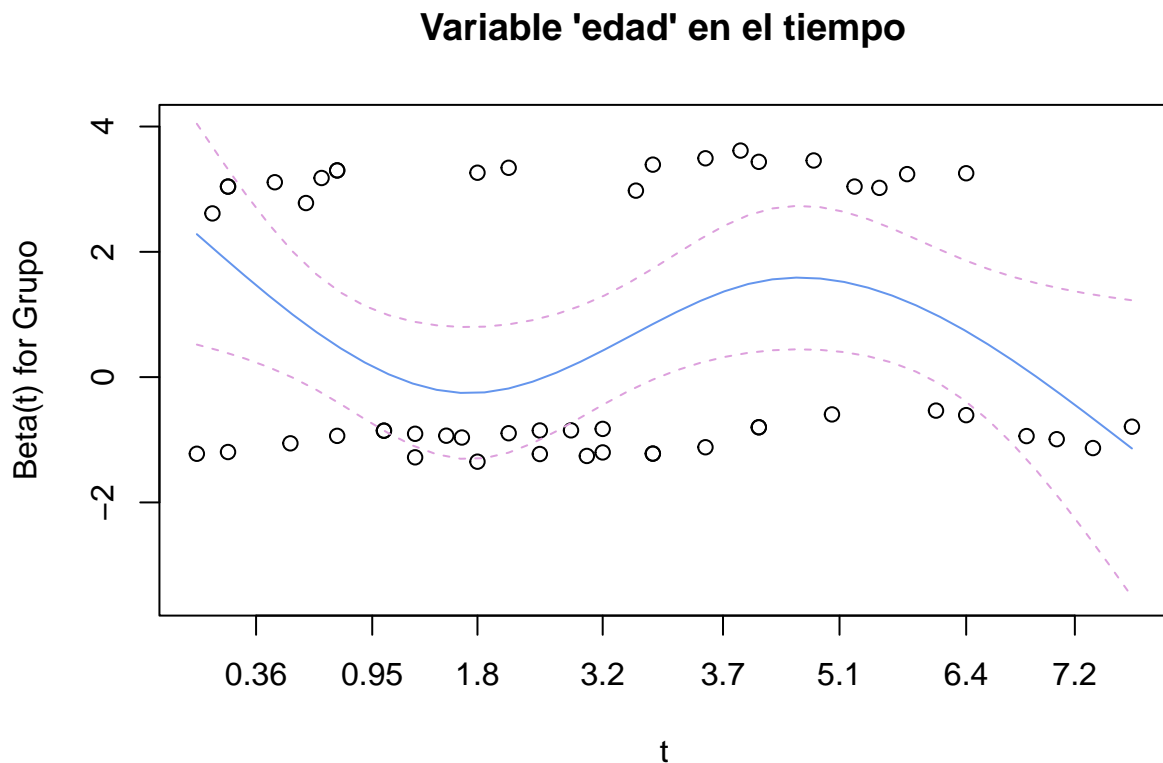
```
##          chisq df          p
```

```
## factor(stage_aux) 4.6536 1 0.031
## Grupo            0.0955 1 0.757
## GLOBAL           4.8091 2 0.090
```

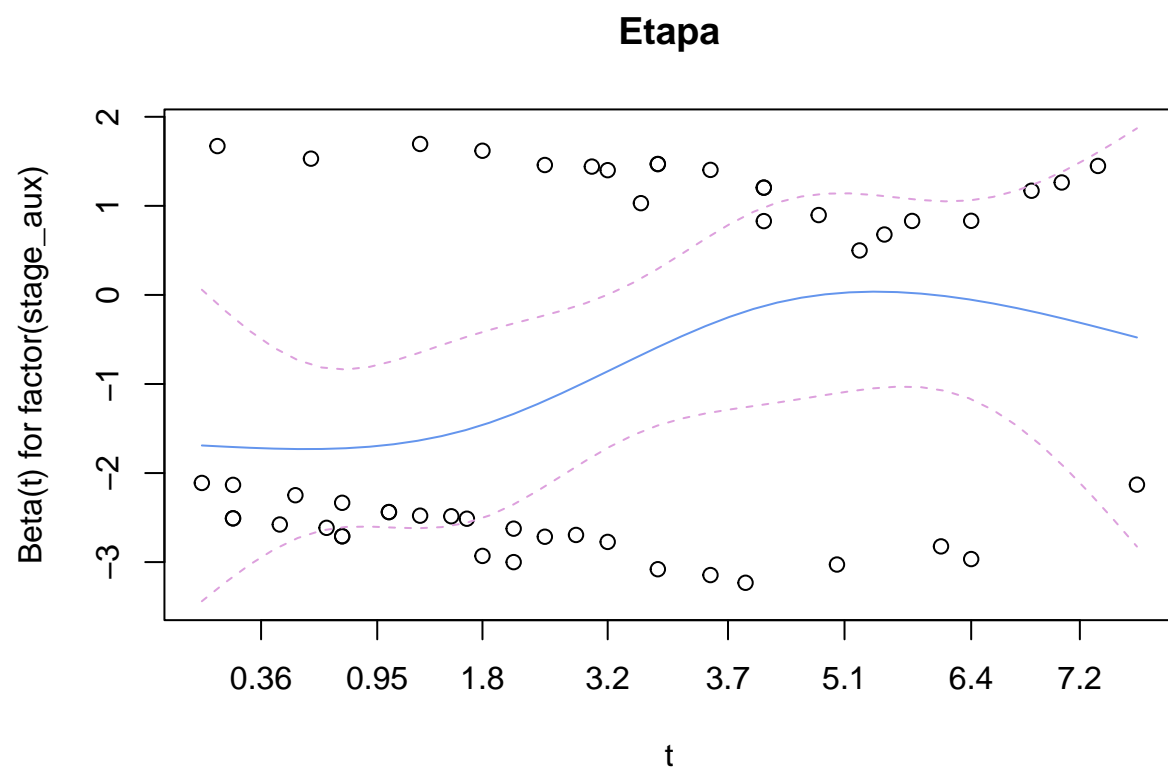
Tenemos el mismo problema, para las etapas no se pasa la prueba, tampoco de manera global, por lo que hay evidencia significativa estadísticamente como para decir que los riesgos no son proporcionales. **Tampoco es válido el supuesto en este modelo**

Verifiquemos gráficamente los residuos Cox - Snell:

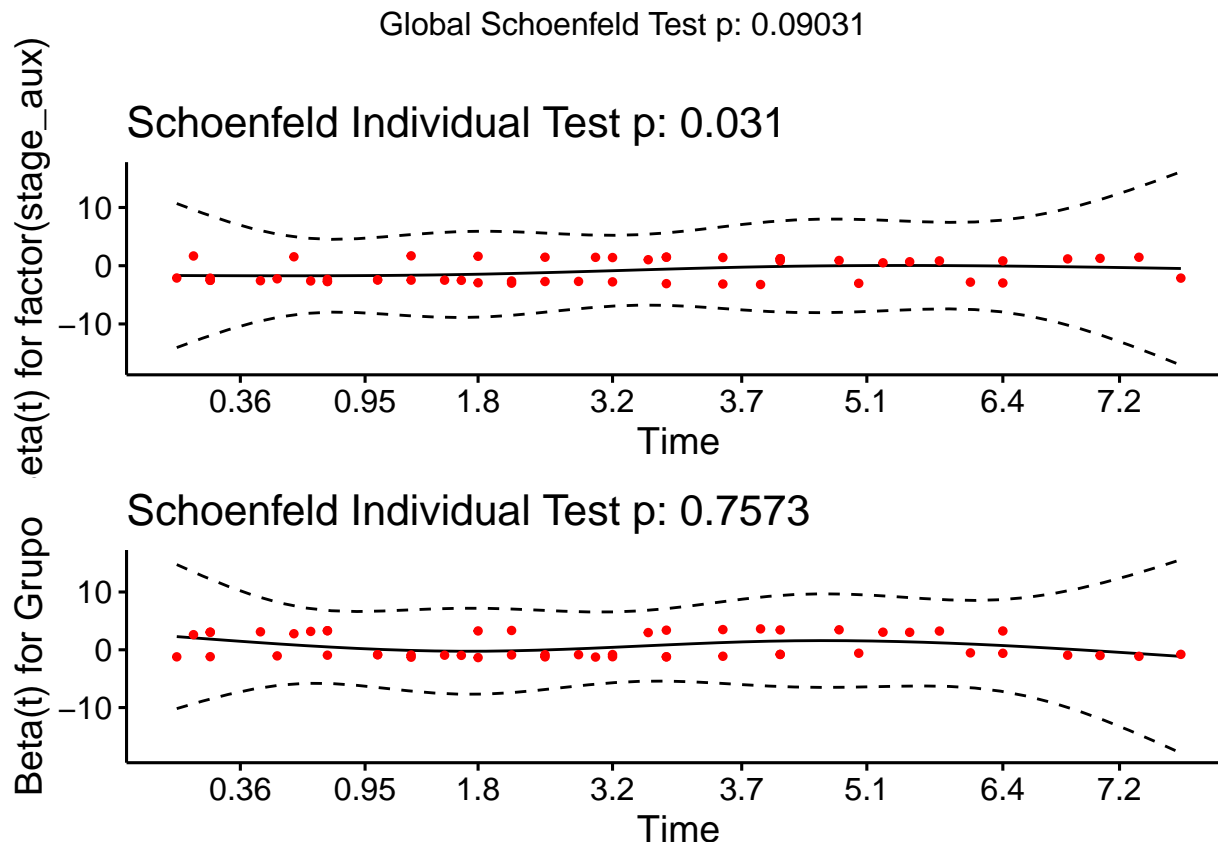
```
plot(sup_3, var = "Grupo", xlab = "t", main = "Variable 'edad' en el tiempo",
      col = c("cornflowerblue", "plum"))
```



```
# Y, finalmente,
plot(sup_3, var = "factor(stage_aux)", xlab = "t",
      main = "Etapa", col = c("cornflowerblue", "plum"))
```



```
ggcoxzph(sup_3)#verificacion grafica
```



Si bien se comporta mejor en la edad, en la etapa seguimos teniendo problemas en la etapa, aunque se comporta mejor en los grupos de edad al tener un p-value más alto, lo que compensa el p-value global, ya que aquí sí se pasa el test de manera conjunta, pero no de manera individual. Inclusive tenemos severos problemas debido a que la línea azul difiere considerablemente del cero.

Aquí no aplican los residuos martingala

Veamos los de devianza:

```
dev_3 = residuals(modelo_conjunto_3, type = "deviance")
min(dev_3)

## [1] -2.548198

max(dev_3)

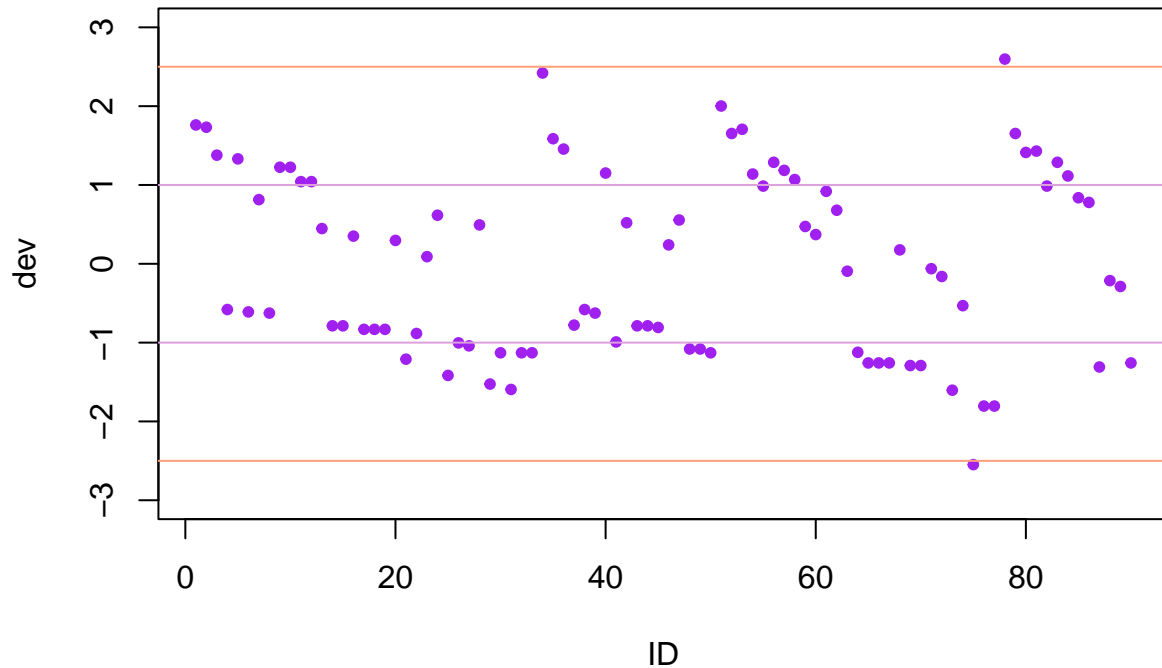
## [1] 2.597437

summary(dev_3)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.54820 -1.03181 -0.07827  0.01970  1.06259  2.59744

plot(dev_3, col = "purple", main = "Residuos de devianza",
      xlab = "ID", ylab = "dev", pch = 20, ylim = c(-3,3))
abline(h = c(-1, 1), col = "plum")
abline(h = c(-2.5, 2.5), col = "lightsalmon")
```

Residuos de devianza



Casos extremos:

`dev_3[abs(dev_3) >= 1] # Posicion`

```
##      1      2      3      5      9     10     11     12
## 1.761923 1.732823 1.378446 1.331439 1.225075 1.225075 1.042703 1.042703
##      21     25     26     27     29     30     31     32
## -1.209900 -1.416323 -1.003310 -1.041308 -1.525992 -1.129258 -1.594117 -1.129258
##      33     34     35     36     40     48     49     50
## -1.129258 2.420594 1.586859 1.455617 1.151367 -1.080998 -1.080998 -1.129258
##      51     52     53     54     56     57     58     64
## 2.002117 1.653224 1.707031 1.138506 1.287799 1.185455 1.069217 -1.122921
##      65     66     67     69     70     73     75     76
## -1.257856 -1.257856 -1.257856 -1.291252 -1.291252 -1.603793 -2.548198 -1.805120
##      77     78     79     80     81     83     84     87
## -1.805120 2.597437 1.653224 1.412801 1.429951 1.287799 1.114514 -1.309371
##      90
## -1.257856
```

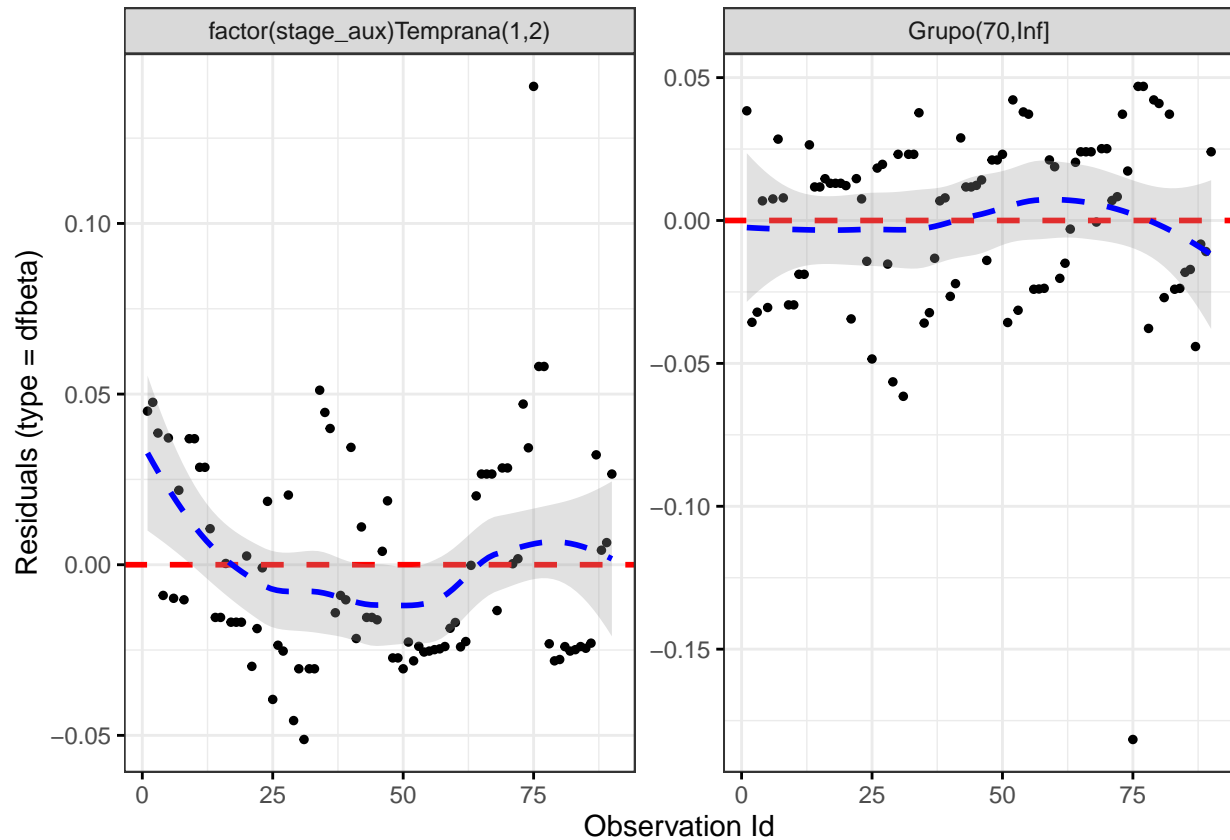
`dev_3[abs(dev_3) >= 2] # Posicion`

```
##      34      51      75      78
## 2.420594 2.002117 -2.548198 2.597437
```

La línea azul se pega bastante al cero, aunque hay varios datos que se salen de las bandas de confianza; aquí tenemos 4 datos que sobrepasan con más de 2 unidades pero es grave ver que hay un dato que difiere casi 2.6 unidades, por lo que están afectando de manera muy considerable nuestro modelo al ser tan atípicos. Además se observa un patrón que decrece, crece y vuelve a crecer. Como líneas diagonales.

```
ggcoxdiagnostics(modelo_conjunto_3, type = "dfbeta")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Aquí la línea azul se pega más al cero, sin embargo seguimos teniendo datos atípicos en la edad, de hecho hay un punto muy, muy extremo que sobresale y puede ser considerado un outlier de bastante importancia que esté afectando considerablemente nuestro modelo.

Modelo ganador

Al pasar el supuesto de riesgos proporcionales en todas sus variables, lo cual consideramos es de mayor peso, creemos que **el modelo 1 es el ganador**, además de que no presenta muchos problemas en el análisis de residuales.